WWU
MÜNSTER

Dissertation

Stability results for advection-diffusion equations with deterministic and random vector fields

Víctor Navarro Fernández

2023

ANGEWANDTE
MATHEMATIK
MÜNSTER

MM
Mathematics
Münster
Cluster of Excellence

Mathematik

Stability results for advection-diffusion equations with deterministic and random vector fields

Inauguraldissertation
zur Erlangung des Doktorgrades der Naturwissenschaften
Doctor rerum naturalium (Dr. rer. nat.)
im Fachbereich Mathematik und Informatik
der Mathematisch-Naturwissenschaftlichen Fakultät
der Westfälischen Wilhelms-Universität Münster

vorgelegt von
Víctor Navarro Fernández
aus Alhama de Murcia

- 2023 -

Dekan: Prof. Dr. Arthur Bartels
Erstgutachter: Prof. Dr. Christian Seis
Zweitgutachter: Prof. Dr. André Schlichting
Drittgutachter: Prof. Dr. Camilla Nobili

Tag der mündlichen Prüfung: 15.01.2024
Tag der Promotion: 15.01.2024

A mi madre, a mi padre y a mi hermano.
Gracias.

Todo pasa y todo queda;
pero lo nuestro es pasar,
pasar haciendo caminos,
caminos sobre la mar.

Antonio Machado

Contents

1. Introduction

In this thesis we address stability results for the transport and advection-diffusion equations of passive scalars driven by deterministic and random vector fields. The study of the dynamics of passive tracers such as temperature, solute concentration, dye or salinity is a topic of great interest to the mathematics and physics community. The transport and advection-diffusion equations describe the evolution of a passive scalar under the action a vector field together with the effect of molecular friction that produces diffusion. These equations model processes that can be found in many different areas, from financial mathematics to the natural sciences such as drift-diffusion processes in semiconductor physics, heat transmission through a fluid layer, turbulence or mixing by stirring on industrials applications.

All the results here presented revolve around the following equation,

$$\partial_t\theta + \nabla\cdot(u\theta) = \kappa\Delta\theta,$$

where θ denotes the passive scalar, u the vector field and $\kappa \geq 0$ the diffusion coefficient. We call this the *transport* equation if there is no diffusion, i.e. when $\kappa = 0$, and the *advection-diffusion* equation in case $\kappa > 0$. In the literature it can be found as well under somewhat different names, e.g. Fokker–Planck or drift-diffusion equation.

Ladyženskaya's classical theory of parabolic equations from the 50s–60s [57, 58] already yields well-posedness for the transport and advection-diffusion equations with smooth coefficients. Since then, many interesting results concerning existence, uniqueness and long-time dynamics have been addressed by the community. The interest has increased since the 80s in questions related to lower regularity assumptions for the vector fields, thus in order to obtain well-posedness for these, new concepts and definitions of what to be a solution means were needed. The notion of distributional, renormalized or Lagrangian solutions were announced for this reason [3, 31, 37, 59], and with new definitions, new doors were opened and more questions were posed.

Even with the simplicity given by the linear structure of the equation, and even taking into account the regularizing effect of the diffusion in case $\kappa > 0$, it has been recently discovered that there are certain vector fields whose regularity is not far from regularities where well-posedness is known, for which there is non uniqueness [69, 70]. Indeed, by means of the method of convex integration, one can construct solutions to the transport and advection-diffusion equations that are trivial everywhere in the domain initially and after some time they stop being trivial.

In this thesis new results are derived regarding stability and mixing estimates for transport and advection-diffusion equations, first for low regularity deterministic vector fields, and second for Lipschitz but random vector fields. We will now give a more detailed overview of the contents included in each chapter of this thesis.

Chapter 2 contains preliminary results which are required for a full understanding of the subsequent chapters. First of all we give a review on well-posedness theory and energy estimates for transport and advection-diffusion equations. Then we introduce some special distances in the space of densities that come from the optimal transport problem and measure convergence in the weak topology. Next we define the notion of Markov process, Markov transition kernels, which are important for the cases in which we study the evolution of passive scalar under the action of stochastic vector fields. We also define the concept of invariant measure and give some classical existence and uniqueness results. Finally we address the mixing problem, since this will be the central object of study in the last chapter. We introduce formal definitions about how to measure the degree of mixedness of a passive scalar and present the idea of enhanced dissipation by incompressible vector fields.

Chapter 3 deals with stability estimates for the advection-diffusion equation with vector fields in low regularity settings. In order to derive such estimates we use distances from the theory of optimal transport. On the one hand we consider vector fields in the DiPerna-Lions setting and derive an estimate that is optimal in some regards that we also discuss. On the other hand we come up with a new uniqueness result for the advection-diffusion equation with vector fields whose gradient is a singular integral of a merely integrable function. This extends the previous work for the transport equation [31] to the diffusive setting.

In Chapter 4 we derive optimal error estimates for an implicit finite volume approximation of the advection-diffusion equation with vector field again in a low regularity setting, namely in the DiPerna-Lions class. By means of both a Eulerian and a Lagrangian perspective, we come up with the appropriate bounds for the distance between the solution from upwind scheme and the exact solution. We use yet again a distance coming from the theory of optimal transport that metrizes weak convergence. This results generalizes the previous work [79, 80] to the diffusive setting, and improves the order of convergence regarding the size of the mesh from 1/2 to 1.

To sum up, Chapter 5 contains results from an unfinished, challenging project regarding ergodicity and mixing properties of randomly driven vector fields. More precisely we derive exponential ergodicity and annealed mixing provided that certain operator is hypocoercive. We introduce a new method to study the mixing problem with stochastic vector fields from a PDE perspective, which extends the previous work based on the Lagrangian picture [7, 8]. In addition, we provide new examples of vector fields that satisfy the required conditions for exponential ergodicity in bounded domain with and without boundary.

Chapter 2 contains no new results. Chapter 3 is based on the article [72], which is a joint work with André Schlichting and Christian Seis. Large parts of it are reproduced verbatim. Chapter 4 is based on the article [71], which is joint with André Schlichting and large parts of it are reproduced verbatim as well. Chapter 5 is a joint work with André Schlichting and Christian Seis and the results included appear for the first time in this thesis.

Acknowledgment

I would like to thank my supervisor Prof. Dr. Christian Seis for his unwavering trust and guidance throughout my PhD. I would like to thank Prof. Dr. André Schlichting as well for his commitment to all the projects here presented. I am very grateful to both for being accessible and open, and for being so patient with my learning process. I am grateful for the many activities they have promoted and organised, in particular the weekly reading seminar *Advanced Topics on Analysis and PDEs* has been really important for my mathematical education.

I wish to say thank you to my workgroup colleagues: Chun Yin Lam, David Meyer, Dominik Winkler, Dr. Lukas Niebel and Stefano Ceci, who have been very supportive and helpful, and from whom I have learned so much. I want to thank as well Prof. Dr. Anna Siffert for her great support as my mentor. I am grateful to all the workers from the Cluster of Excellence and the Institut für Analysis und Numerik who have created an excellent atmosphere with great opportunities.

I wish to say thanks to the many people I met in conferences, workshops and summer schools, for introducing me to new problems and perspectives in research. In particular I would like to thank Dr. Rishabh Gvalani for inviting me to the Max–Planck–Insitut in Leipzig, and for providing important insights on my work.

I am very grateful to all my dear colleagues and friends from the department of mathematics –including Mina and Sisqo– for the long discussions, the coffee breaks, the very many cakes, and the walks around the Aasee. They have created all over the best environment for work and off-work, even through a pandemic and lockdowns. Further, I would also like to thank all my beloved friends in the Studentenorchester Münster and the HSP Leichtathletik group for their kindness and hospitality.

Last but not least, I owe a wholeheartedly thank you to my family, especially my mother, my father and my brother, who have constantly been the best support, and have provided an inestimable safety net.

I gratefully acknowledge the financial support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2044 –390685587, Mathematics Münster: Dynamics–Geometry–Structure and by the DFG Grant 432402380.

2. Preliminary results

This chapter contains some relevant definitions, concepts, and preliminary results that will be needed for the subsequent chapters of this thesis. There are no new results included.

On a first note let us define some symbols, abbreviations, notation and conventions that we will use along the monograph.

- D is a finite dimensional Polish space. In most cases we will consider D to be $\mathbb{R}^d$ or a bounded subset of $\mathbb{R}^d$, such as the torus $\mathrm{D} = \mathbb{T}^d = \mathbb{R}^d/[0, 2\pi)$ or a bounded domain with a sufficiently regular boundary.

- Given a subset $A \subseteq \mathrm{D}$, we denote its complementary as $A^c = \mathrm{D} \setminus A$

- Given a function $f : \mathrm{D} \to \mathbb{R}$ we write $f^+ = \max\{f, 0\}$ for the positive part of f, and $f^- = -\min\{f, 0\}$ for the negative part so that $f = f^+ - f^-$ and $|f| = f^+ + f^-$.

- $\mathcal{B}(\mathrm{D})$ is the Borel $\sigma-$algebra of D.

- $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, where Ω is the sample space, $\mathcal{F}$ a $\sigma-$algebra and $\mathbb{P}$ a probability measure. A typical example that we use is the probability space of Brownian motion, where $\Omega = C_0([0, \infty), \mathbb{R}^d)$ is the set of all continuous path with $\omega(0) = 0$, $\mathcal{F} = \{\mathcal{F}_t\}_{t\geq 0}$ is a filtered $\sigma-$algebra and $\mathbb{P}$ a Wiener probability measure.

- $\mathcal{P}(\mathrm{D})$ is the set of all probability measures in D.

- Given a bijective map $T : \mathrm{D} \to \mathrm{D}$ and a measure μ on D, we define the push-forward measure $T_{\#}\mu$ as

  $$T_{\#}\mu(A) = \mu(T^{-1}A) \quad \text{for all } A \in \mathcal{B}(\mathrm{D}).$$

  If the measure μ admits a density $f : \mathrm{D} \to \mathbb{R}$ such that $\mathrm{d}\mu(x) = f(x)\,\mathrm{d}x$, the push-forward measure can be defined via

  $$\int_{\mathrm{D}} \phi(x)(T_{\#}f)(x)\,\mathrm{d}x = \int_{\mathrm{D}} \phi(T(x))f(x)\,\mathrm{d}x,$$

  for all ϕ sufficiently regular.

- Unless otherwise is stated, $\|\cdot\|$ denotes the $L^2(\mathrm{D})$ norm, and we write $\langle\cdot,\cdot\rangle$ for the standard scalar product in $L^2(\mathrm{D})$,

$$\langle f, g\rangle = \int_{\mathrm{D}} f g \, \mathrm{d}x,$$

namely $\|f\|^2 = \langle f, f\rangle$.

- We use the Bochner space notation $L^r(L^s)$ to denote the space $L^r((0,T);L^s(\mathrm{D}))$ and similarly for other Banach spaces.

- We write $a \lesssim b$ if there exists a constant $C > 0$ such that $a \leq Cb$.

2.1. Energy estimates for advection-diffusion equations

Let $\mathrm{D} \subseteq \mathbb{R}^d$ be a domain with or without boundary and $0 < T \leq +\infty$ a maximal time of existence. Consider the advection-diffusion equation that models the evolution of a passive scalar $\theta : (0,T) \times \mathrm{D} \to \mathbb{R}$ driven by a vector field $u : (0,T) \times \mathrm{D} \to \mathbb{R}^d$ given by the Cauchy problem,

$$\begin{cases} \partial_t\theta + \nabla\cdot(u\theta) &= \kappa\Delta\theta & \text{in } (0,T)\times\mathrm{D}, \\ n\cdot(u-\kappa\nabla\theta) &= 0 & \text{on } (0,T)\times\partial\mathrm{D}, \\ \theta(0,\cdot) &= \theta^0 & \text{in } \mathrm{D}. \end{cases} \tag{2.1}$$

Here $\kappa \geq 0$ is the diffusion coefficient $n(x)$ is the outer unit vector at position $x \in \partial\mathrm{D}$. If $\mathrm{D} = \mathbb{R}^d$ or $\mathrm{D} = \mathbb{T}^d$, then the boundary conditions disappear from the equation.

One straightforward consequence of this definition is that, due to linearity, the mass of the passive scalar is conserved in D for all times, namely

$$\int_{\mathrm{D}} \theta(t,\cdot)\,\mathrm{d}x = \int_{\mathrm{D}} \theta^0\,\mathrm{d}x \quad \text{for all } t \in (0,T).$$

Well-posedness of solutions for smooth vector fields and initial data goes back to the classical theory of parabolic equations, see Ladyženskaya et al. [58]. In some specific contexts in physics, for instance, when studying the transport of a mass, dye, or any scalar quantity by a turbulent flow [74, 86], the vector field involved has a very low regularity, thus a mathematical theory for transport and advection-diffusion equations with rough vector fields is needed. In this regard, we define the following notion of solution to (2.1).

Definition 2.1. Let $u \in L^1((0,T); L^p_{\mathrm{loc}}(\mathrm{D}))$ and $\theta^0 \in L^q_{\mathrm{loc}}(\mathrm{D})$ be given for some

$$p, q \in [1, +\infty] \quad \text{with} \quad \frac{1}{p} + \frac{1}{q} \leq 1.$$

A function $\theta \in L^\infty((0,T); L^q_{\mathrm{loc}}(\mathrm{D}))$ is called a *distributional solution* of (2.1) if

$$\iint_{(0,T)\times\mathrm{D}} \theta \, (\partial_t \phi + u \cdot \nabla\phi + \kappa \Delta\phi) \, \mathrm{d}x \, \mathrm{d}t + \int_{\mathrm{D}} \theta^0 \, \phi|_{t=0} \, \mathrm{d}x = 0$$

holds for all $\phi \in C^\infty_c([0,T) \times \mathrm{D})$.

In order to obtain uniqueness of distributional solutions, a customary aproach is based on energy estimates. Testing with the appropriate functions in (2.1) we obtain the following, at first formal, energy estimates:

$$\frac{1}{q(q-1)} \frac{d}{dt} \int_{\mathrm{D}} |\theta|^q \, \mathrm{d}x + \kappa \int_{\mathrm{D}} |\theta|^{q-2} |\nabla\theta|^2 \, \mathrm{d}x \leq \int_{\mathrm{D}} |\theta|^{q-1} |\nabla\theta| |u| \, \mathrm{d}x, \tag{2.2}$$

$$\frac{1}{q(q-1)} \frac{d}{dt} \int_{\mathrm{D}} |\theta|^q \, \mathrm{d}x + \kappa \int_{\mathrm{D}} |\theta|^{q-2} |\nabla\theta|^2 \, \mathrm{d}x \leq \int_{\mathrm{D}} |\theta|^q (\nabla \cdot u)^- \, \mathrm{d}x, \tag{2.3}$$

for any $q > 1$. The limit $q \to 1$ leads to an apriori estimate in terms of the entropy of the form,

$$\frac{d}{dt} \int_{\mathrm{D}} \theta \log\theta \, \mathrm{d}x + \kappa \int_{\mathrm{D}} \frac{|\nabla\theta|^2}{\theta} \, \mathrm{d}x \leq \int_{\mathrm{D}} |\theta| (\nabla \cdot u)^- \, \mathrm{d}x. \tag{2.4}$$

These energy estimates provide a framework to deal with well-posedness of equation (2.1) assuming different regularities for the vector field. In particular we are mostly interested in the following settings.

1. *The Ladyženskaya–Prodi–Serrin setting*. Estimate (2.2) yields that there exists a unique distributional solution to (2.1)

   $$\theta \in L^\infty((0,T); L^2(\mathrm{D})),$$

   provided that $\theta^0 \in L^2(\mathrm{D})$ and $u \in L^r((0,T); L^p(\mathrm{D}))$ with

   $$\frac{2}{r} + \frac{d}{p} \leq 1, \quad \text{with} \quad \begin{array}{l} r \in [2,\infty) \text{ and } p \in (d,\infty] \text{ if } d \geq 2, \\ r \in [2,4] \text{ and } p \in [2,\infty] \text{ if } d = 1. \end{array}$$

   This setting has been recently revisited in [11].

2. *The DiPerna–Lions setting*. Estimate (2.3) yields that there exists a unique

distributional solution to (2.1)

$$\theta \in L^\infty((0,T); L^q(\mathrm{D})),$$

provided that $\theta^0 \in L^q(\mathrm{D})$ and the vector field is *weakly compressible*, namely

$$u \in L^1((0,T); W^{1,p}(\mathrm{D})) \quad \text{and} \quad (\nabla \cdot u)^- \in L^1((0,T); L^\infty(\mathrm{D}))$$

with $p, q \geq 1$ Hölder conjugate or larger,

$$\frac{1}{p} + \frac{1}{q} \leq 1.$$

This is the most relevant regularity setting for Chapters 3 and 4. This result can be found in [37, 59].

In general, if $\kappa > 0$ we also obtain a control over the derivatives of the passive scalar. For instance, if D is bounded or $\theta^0 \in L^1 \cap L^q$ has finite first moments, we obtain that

$$\nabla\theta \in L^1((0,T); L^1(\mathrm{D})),$$

which is of relevance for Chapter 3.

The DiPerna–Lions theory was further extended by Ambrosio [3] to weakly compressible vector fields of $BV-$regularity. In addition, in Chapter 3, we address the problem with vector fields whose gradient is a singular integral of an L^1 function [31, 72], which is of interest for certain problems in fluid dynamics.

The condition for the coefficients given by the DiPerna–Lions setting is optimal for *both* $\kappa > 0$ and $\kappa = 0$. Indeed, with the method of convex integration, Modena, Sattig, and Székelyhidi [69, 70] showed that there are Sobolev vector fields

$$u \in C([0,T]; (L^p \cap W^{1,p'})(\mathbb{T}^d))$$

for which uniqueness fails in the class of densities $\theta \in C([0,T]; L^q(\mathbb{T}^d))$ with $p, q \in (1, \infty)$, $p' \in [1, \infty)$ such that there holds

$$\frac{1}{p} + \frac{1}{q} = 1, \quad \frac{1}{p'} + \frac{1}{q} > 1 + \frac{1}{d} \quad \text{and} \quad p < d.$$

An improvement on the above condition is obtained by Cheskidov and Luo [22] for the transport equation ($\kappa = 0$) at the expense of a worse time-integrability, that is solutions

$\theta \in L^1([0,T];L^q(\mathbb{T}^d))$ with

$$\frac{1}{p}+\frac{1}{q}>1.$$

2.2. Optimal transport distances

In this section we introduce some tools from the theory of optimal transportation that will be useful in the different stability results from theorems in Chapters 3 and 4. We decide to make a presentation suitable for our needs, in a rather smooth setting that is enough for our purposes. For generalizations and detailed proofs of the subsequent results, we refer to Villani's monograph [94].

Definition 2.2. Given two nonnegative densities

$$\mu_1,\mu_2 \in L^1_+(\mathbb{R}^d)=\{\mu\in L^1(\mathbb{R}^d):\mu\geq 0\},$$

we define $\Pi(\mu_1,\mu_2)$ to be the set of all transport plans between μ_1 and μ_2. Namely, $\pi\in\Pi(\mu_1,\mu_2)$ is a measure in $\mathbb{R}^d\times\mathbb{R}^d$ with the property

$$\pi[A\times\mathbb{R}^d]=\mu_1[A],\quad \pi[\mathbb{R}^d\times A]=\mu_2[A],\quad \text{for all } A\in\mathcal{B}(\mathbb{R}^d).$$

Analogously, it can also be define via the integral formulation, namely we say $\pi\in\Pi(\mu_1,\mu_2)$ if there holds

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} f(x_1)\,\mathrm{d}\pi(x_1,x_2)=\int_{\mathbb{R}^d} f(x_1)\,\mathrm{d}\mu_1(x_1),$$
$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} f(x_2)\,\mathrm{d}\pi(x_1,x_2)=\int_{\mathbb{R}^d} f(x_2)\,\mathrm{d}\mu_2(x_2).$$

We want to consider distances arising from the optimal transportation problem. We say that a function

$$c:[0,\infty)\to[0,\infty)$$

is a *cost function* if it continuous and nondecreasing. The *optimal transport problem* or also called *Kantorovich problem*, is defined as a minimization problem of the total transportation cost from one configuration μ_1 to another configuration μ_2. We denote this quantity by

$$\mathcal{D}_c(\mu_1,\mu_2)=\inf_{\pi\in\Pi(\mu,\nu)}\iint_{\mathbb{R}^d\times\mathbb{R}^d} c(|x-y|)\,\mathrm{d}\pi(x,y). \tag{2.5}$$

From a physical perspective one could say that $\mathcal{D}_c(\mu_1,\mu_2)$ measures the minimal total cost of transporting an initial configuration of mass or goods given by μ to a final configuration

ν if the cost of the transport of an infinitesimal part is modelled by c.

For our purposes we want to consider cost functions $c : [0,\infty) \to [0,\infty)$ that are strictly concave, Lipschitz with uniform Lipschitz constant L_c and such that $c(0) = 0$. This type of cost functions induces a metric

$$d(x,y) = c(|x-y|) \quad \text{on } \mathbb{R}^d.$$

Moreover, as it is proved in [94, Theorem 1.14], in this case the optimal transportation problem (2.5) admits a dual formulation,

$$\mathcal{D}_c(\mu_1,\mu_2) = \sup_{|\zeta(x)-\zeta(y)|\le c(|x-y|)} \int_{\mathbb{R}^d} (\mu_1(x)-\mu_2(x))\zeta(x)\,\mathrm{d}x. \tag{2.6}$$

Note then that $\mathcal{D}_c(\mu_1,\mu_2)$ is a transshipment cost that only depends on the difference $\mu_1-\mu_2$. This allows us to extend the theory to densities that are not necessarily nonnegative but just that verify $\mu[\mathbb{R}^d] = \nu[\mathbb{R}^d] \in \mathbb{R}$.

In addition, because $d(x,y) = c(|x-y|)$ is a metric on $\mathbb{R}^d$, $\mathcal{D}_c(\cdot,\cdot)$ defines a metric on the space of densities with the same total mass and it is usually referred to as the *Kantorovich–Rubinstein distance* or, *optimal transportation distance*. Therefore, for any function $\theta \in L^1(\mathbb{R}^d)$ with zero total mass, $\theta^+[\mathbb{R}^d] = \theta^-[\mathbb{R}^d]$, we conveniently define the norm

$$\mathcal{D}_c(\theta) = \mathcal{D}_c(\theta^+,\theta^-). \tag{2.7}$$

It is known that the first problem (2.5) admits a unique minimizer, in general $\pi_{\text{opt}} \in \Pi(\mu,\nu)$, named the optimal transport plan. The dual problem (2.6) also admits a maximizer, that could be nonunique ζ_{opt}, called the Kantorovich potential. It is characterized by the relation

$$\zeta_{\text{opt}}(x) - \zeta_{\text{opt}}(y) = c(|x-y|) \qquad \text{for } d\pi_{\text{opt}} - \text{almost all } (x,y).$$

We can weakly-differentiate this identity to obtain

$$\nabla_x\zeta_{\text{opt}}(x) = \nabla_y\zeta_{\text{opt}}(y) = c'(|x-y|)\frac{x-y}{|x-y|} \qquad \text{for } d\pi_{\text{opt}} - \text{almost all } (x,y),$$

and therefore it holds $|\nabla\zeta_{\text{opt}}| \le L_c$, since the cost function is L_c−Lipschitz.

There is nonetheless an additional way of presenting the optimal transport plan $\pi_{\text{opt}} \in \Pi(\mu,\nu)$ when the cost function is strictly concave. Gangbo and McCann [48] proved that there exist measurable maps $T, S : \mathbb{R}^d \to \mathbb{R}^d$ such that there holds

$$\pi_{\text{opt}} = (\text{id} \times T)_{\#}\mu_1 = (S \times \text{id})_{\#}\mu_2$$

and where T and S obey the relations $\mu = S_{\#}\nu$, $\nu = T_{\#}\mu$. This characterization will be useful in the subsequent sections to prove some relevant results.

The following are examples of optimal transport distances that are relevant for the subsequent chapters of this thesis.

- $1-$ *Wasserstein distance* W_1: This distance is defined from the cost function $c(z) = z$, and using the dual representation (2.6) it can be represented in terms of $1-$Lipschitz functions as

$$W_1(\mu_1, \mu_2) = \sup_{\|\nabla\zeta\|_{L^\infty} \leq 1} \int_{\mathbb{R}^d} (\mu_1(x) - \mu_2(x))\zeta(x)\,\mathrm{d}x. \tag{2.8}$$

  One can see from this representation that for mean zero functions θ, the norm $W_1(\theta^+, \theta^-)$ coincides with the norm of the negative Sobolev space $\dot{W}^{-1,\infty}(\mathbb{R}^d)$. The Wasserstein distances are in general defined for any $p \geq 1$ through the convex cost function $c(z) = c^p$.

- *Logarithmic optimal transport distance* $\mathcal{D}_\delta$: This is the distance we use for most of the results presented in Chapters 3 and 4. It is defined, for any given $\delta > 0$ by the cost function

$$c(z) = \log\left(\frac{z}{\delta} + 1\right). \tag{2.9}$$

  This is a concave cost that produces the following bound for the optimal Kantorovich potential,

$$\|\nabla\zeta_{\text{opt}}\|_{L^\infty} \leq \frac{1}{\delta}.$$

- *Bounded distance* $\mathcal{D}^b$: This distance function is defined via the cost function,

$$c(z) = \tanh(z). \tag{2.10}$$

  Compared to the logarithmic case, it has the property of being bounded as the name indicates, which comes in handy for some applications.

Recall that by (2.7) both $\mathcal{D}_\delta$ and $\mathcal{D}^b$ induce norms in the space of zero average densities. For every $\theta \in L^1(\mathbb{R}^d)$ mean zero, $\mathcal{D}^b(\theta)$ can be controlled by $\mathcal{D}_\delta(\theta)$ through the following Lemma, introduced and first proved in [83] and later adapted in [31] to a more convenient framework for our purposes here.

Lemma 2.1. *Let $\theta \in L^1(\mathbb{R}^d)$ and mean zero. Then $\forall \gamma, \delta > 0$ it holds*

$$\mathcal{D}^b(\theta) \leq \frac{\mathcal{D}_\delta(\theta)}{\log \frac{1}{\gamma}} + \frac{\delta}{\gamma}\|\theta\|_{L^1}.$$

A nice and short proof of this Lemma can be found in [31, Lemma 3.1].

In the forthcoming sections we will apply these optimal transportation distances with densities that depend not only on $x \in \mathbb{R}^d$ but also on $t \in (0,T)$. Therefore the optimal transport plans or the Kantorovich potentials might be time-dependent. In order to simplify the notation we will refer as π_t to the optimal transport plan π_{opt} associated to the distance $\mathcal{D}_c(\mu_1(t,\cdot), \mu_2(t,\cdot))$, with $t \in (0,T)$. Analogously we write ζ_t to denote the Kantorovich potential ζ_{opt} associated to the same distance.

In order to obtain stability for optimal transport distances, we need to introduce the following key result. It was first proved in [83], however we include the proof from [72] here because we consider that it might be clarifying for the reader.

Lemma 2.2. *Let μ_1 and μ_2 be two distributional solutions in $L^1(W^{1,1})$ to the advection-diffusion equation with advection fields u_1, u_2 and diffusion coefficients $\kappa_1, \kappa_2 > 0$ respectively. Then the mapping $t \mapsto \mathcal{D}_c(\mu_1(t,\cdot), \mu_2(t,\cdot))$ is absolutely continuous with*

$$\begin{aligned}\frac{d}{dt}\mathcal{D}_c(\mu_1(t,\cdot), \mu_2(t,\cdot)) = &\int_{\mathbb{R}^d} \nabla\zeta_t \cdot (u_1(t,x)\mu_1(t,x) - u_2(t,x)\mu_2(t,x))\,\mathrm{d}x \\ &- \int_{\mathbb{R}^d} \nabla\zeta_t \cdot (\kappa_1\nabla\mu_1(t,x) - \kappa_2\nabla\mu_2(t,x))\,\mathrm{d}x\end{aligned} \tag{2.11}$$

where ζ_t is the Kantorovich potential corresponding to $\mathcal{D}_c(\mu_1(t,\cdot), \mu_2(t,\cdot))$.

Proof. By the definition of distributional solution, by a standard approximation argument and integrating by parts, for any $h \in \mathbb{R}$ such that $t - h \in (0,T)$ we have that

$$\begin{aligned}\int_{\mathbb{R}^d} \zeta(\mu_i(t,x) - \mu_i(t-h,x))\,\mathrm{d}x &= \int_{\mathbb{R}^d} \nabla\zeta \cdot \int_{t-h}^t u_i(s,x)\mu_i(s,x)\,\mathrm{d}s\,\mathrm{d}x \qquad (2.12)\\ &\quad + \kappa_i \int_{\mathbb{R}^d} \zeta \int_{t-h}^t \Delta\mu_i(s,x)\,\mathrm{d}s\,\mathrm{d}x \\ &= \int_{\mathbb{R}^d} \nabla\zeta \cdot \int_{t-h}^t u_i(s,x)\mu_i(s,x)\,\mathrm{d}s\,\mathrm{d}x \\ &\quad - \kappa_i \int_{\mathbb{R}^d} \nabla\zeta \cdot \int_{t-h}^t \nabla\mu_i(s,x)\,\mathrm{d}s\,\mathrm{d}x \qquad (2.13)\end{aligned}$$

for all $\zeta \in C_c^\infty(\mathbb{R}^d)$, almost every $t \in (0,T)$ and both $i \in \{1,2\}$. Since now μ_i and $u_i\mu_i$ are in $L^1(L^1)$ we are allowed to consider (2.13) for all $\zeta \in W^{1,1}(\mathbb{R}^d)$.

First we will show that the mapping $t \mapsto \mathcal{D}_c(\mu_1(t,\cdot),\mu_2(t,\cdot))$ is absolutely continuous and therefore differentiable almost everywhere in $(0,T)$. By the optimality of the Kantorovich potential ζ_t at time t it holds for almost every $t \in (0,T)$ that

$$\begin{aligned}
&\mathcal{D}_c(\mu_1(t,\cdot),\mu_2(t,\cdot)) - \mathcal{D}_c(\mu_1(t-h,\cdot),\mu_2(t-h,\cdot)) \\
&\quad \le \int_{\mathbb{R}^d} \zeta_t(\mu_1(t,x)-\mu_1(t-h,x))\,\mathrm{d}x - \int_{\mathbb{R}^d} \zeta_t(\mu_2(t,x)-\mu_2(t-h,x))\,\mathrm{d}x \\
&\quad = \int_{\mathbb{R}^d} \nabla\zeta_t \cdot \int_{t-h}^t u_1(s,x)\mu_1(s,x)\,\mathrm{d}s\,\mathrm{d}x - \int_{\mathbb{R}^d} \nabla\zeta_t \cdot \int_{t-h}^t u_2(s,x)\mu_2(s,x)\,\mathrm{d}s\,\mathrm{d}x \\
&\qquad - \kappa_1 \int_{\mathbb{R}^d} \nabla\zeta_t \cdot \int_{t-h}^t \nabla\mu_1(s,x)\,\mathrm{d}s\,\mathrm{d}x + \kappa_2 \int_{\mathbb{R}^d} \nabla\zeta_t \cdot \int_{t-h}^t \nabla\mu_2(s,x)\,\mathrm{d}s\,\mathrm{d}x. \qquad (2.14)
\end{aligned}$$

Analogously, again by the optimality of the Kantorovich potential ζ_{t-h} at time $t-h$, it holds for almost every $t \in (0,T)$ that

$$\begin{aligned}
&\mathcal{D}_c(\mu_1(t,\cdot),\mu_2(t,\cdot)) - \mathcal{D}_c(\mu_1(t-h,\cdot),\mu_2(t-h,\cdot)) \\
&\quad \le \int_{\mathbb{R}^d} \zeta_{t-h}(\mu_1(t,x)-\mu_1(t-h,x))\,\mathrm{d}x - \int_{\mathbb{R}^d} \zeta_{t-h}(\mu_2(t,x)-\mu_2(t-h,x))\,\mathrm{d}x \\
&\quad = \int_{\mathbb{R}^d} \nabla\zeta_{t-h} \cdot \int_{t-h}^t u_1(s,x)\mu_1(s,x)\,\mathrm{d}s\,\mathrm{d}x - \int_{\mathbb{R}^d} \nabla\zeta_{t-h} \cdot \int_{t-h}^t u_2(s,x)\mu_2(s,x)\,\mathrm{d}s\,\mathrm{d}x \\
&\qquad - \kappa_1 \int_{\mathbb{R}^d} \nabla\zeta_{t-h} \cdot \int_{t-h}^t \nabla\mu_1(s,x)\,\mathrm{d}s\,\mathrm{d}x + \kappa_2 \int_{\mathbb{R}^d} \nabla\zeta_{t-h} \cdot \int_{t-h}^t \nabla\mu_2(s,x)\,\mathrm{d}s\,\mathrm{d}x. \qquad (2.15)
\end{aligned}$$

Therefore, using that ζ_t is Lipschitz with Lipschitz constant uniformly bounded by L_c, we can combine (2.14) and (2.15) to obtain

$$\begin{aligned}
&|\mathcal{D}_c(\mu_1(t,\cdot),\mu_2(t,\cdot)) - \mathcal{D}_c(\mu_1(t-h,\cdot),\mu_2(t-h,\cdot))| \\
&\quad \le L_c \int_{t-h}^t \int_{\mathbb{R}^d} |u_1(s,x)\mu_1(s,x)|\,\mathrm{d}s\,\mathrm{d}x + L_c \int_{t-h}^t \int_{\mathbb{R}^d} |u_2(s,x)\mu_2(s,x)|\,\mathrm{d}s\,\mathrm{d}x \\
&\qquad + L_c\kappa_1 \int_{t-h}^t \int_{\mathbb{R}^d} |\nabla\mu_1(s,x)|\,\mathrm{d}s\,\mathrm{d}x + L_c\kappa_2 \int_{t-h}^t \int_{\mathbb{R}^d} |\nabla\mu_2(s,x)|\,\mathrm{d}s\,\mathrm{d}x
\end{aligned}$$

for almost every $t \in (0,T)$. Since $u_i\mu_i \in L^1(L^1)$ and $\nabla\mu_i \in L^1(L^1)$ for $i \in \{1,2\}$, we conclude that indeed $t \mapsto \mathcal{D}_c(\theta(t,\cdot))$ is an absolutely continuous mapping.

It remains to prove that the derivative of the mapping takes the expression (2.11). In order to do that, we consider again (2.14) and (2.15), divide by h and let $h \to 0$. By

Lebesgue's differentiation theorem we get

$$\begin{aligned}\lim_{h\to 0^+} &\frac{\mathcal{D}_c(\mu_1(t,\cdot),\mu_2(t,\cdot)) - \mathcal{D}_c(\mu_1(t-h,\cdot),\mu_2(t-h,\cdot))}{h}\\ &\le \int_{\mathbb{R}^d} \nabla\zeta_t \cdot (u_1(t,x)\mu_1(t,x) - u_2(t,x)\mu_2(t,x))\,\mathrm{d}x\\ &\quad - \int_{\mathbb{R}^d} \nabla\zeta_t \cdot (\kappa_1\nabla\mu_1(t,x) - \kappa_2\nabla\mu_2(t,x))\,\mathrm{d}x\end{aligned}$$

and

$$\begin{aligned}\lim_{h\to 0^-} &\frac{\mathcal{D}_c(\mu_1(t,\cdot),\mu_2(t,\cdot)) - \mathcal{D}_c(\mu_1(t-h,\cdot),\mu_2(t-h,\cdot))}{h}\\ &\ge \int_{\mathbb{R}^d} \nabla\zeta_t \cdot (u_1(t,x)\mu_1(t,x) - u_2(t,x)\mu_2(t,x))\; dx\\ &\quad - \int_{\mathbb{R}^d} \nabla\zeta_t \cdot (\kappa_1\nabla\mu_1(t,x) - \kappa_2\nabla\mu_2(t,x))\,\mathrm{d}x\end{aligned}$$

which implies (2.11) for almost every $t \in (0,T)$. ■

2.3. Markov processes and invariant measures

In Chapter 5 we consider some solutions to the transport and advection-diffusion equation with random vector fields. In order to get a full understanding of these objects, especially when dealing with the Lagrangian perspective, it is important to introduce some Markov processes generated by the stochastic flow.

Here we present a summary of some definitions and properties of Markov processes that will be relevant for our purposes later. For details and further information about the theory of Markov processes and their invariant measures see [32, 50, 51].

Consider a probability space $(\Omega,\mathcal{F},\mathbb{P})$ and a filtration of the $\sigma-$algebra $\{\mathcal{F}_t\}_{t\ge 0}$. Let $X_t : \Omega\times \mathrm{D} \to \mathrm{D}$ be a stochastic flow in $\mathrm{D} \subseteq \mathbb{R}^d$ with $t \ge 0$ that is adapted to the filtration, namely X_t is a $(\mathcal{F}_t \otimes \mathcal{B}(\mathrm{D}),\mathcal{B}(\mathrm{D}))-$measurable function for each $t \ge 0$.

Definition 2.3. We say that the stochastic process $\{X_t\}_{t\ge 0}$ has the Markov property if for every $A \in \mathcal{B}(\mathrm{D})$ and all $0 \le s < t$ there holds

$$\mathbb{P}[X_t \in A \mid \mathcal{F}_s] = \mathbb{P}[X_t \in A \mid X_s].$$

Markov processes can be understood through their *transition kernels*. In our continuous time situation we will denote the Markov kernel of the process $\{X_t\}_{t\geq 0}$ by

$$P_t(x, A) = \mathbb{P}[X_t \in A \mid X_0 = x],$$

with $t > 0$, $A \in \mathcal{F}$ and $x \in \mathrm{D}$. This object, in general words, represents the probability for the stochastic flow X_t to be in the Borel set $A \subset \mathrm{D}$ at time $t > 0$ provided that initially it was at position $x \in \mathrm{D}$.

Markov transition kernels $\{P_t\}_{t\geq 0}$ can act on different type of objects. Next in order we introduce some relevant properties of the Markov kernels since these are of great importance when studying the dynamics of the stochactic flows.

- $P_t(x, \cdot)$ is a Borel probability measure on D.
- Markov kernels act on continuous functions $\phi \in C(\mathrm{D})$,

$$P_t\phi(x) = \int_{\mathrm{D}} \phi(y) P_t(x, \mathrm{d}y).$$

- Regarded as operators $P_t : C \to C$, Markov kernels have the semigroup property,

$$P_0 = \mathrm{id}, \quad P_{t+s} = P_t \circ P_s \quad \text{for all } t, s \geq 0.$$

  In this way we can refer to $\{P_t\}_{t\geq 0}$ as the *Markov semigroup*.

Definition 2.4. Let $\{P_t\}_{t\geq 0}$ be a Markov transition kernel in D as defined before and let $C_b = C \cap L^\infty$ be the space of continuous and bounded functions.

1. We say that the Markov process has the *Feller property* if $P_t\phi \in C_b$ for all $\phi \in C_b$.
2. We say that $\{P_t\}_{t\geq 0}$ has the *strong Feller property* if $P_t\phi \in C_b$ for any continuous function $\phi \in C$ that is not necessarily bounded.

Notice that if the mapping $x \mapsto P_t(x, \cdot)$ is weak* continuous on the space of probability measures in D, then P_t is Feller. In relation to this, it makes sense to consider the action of the Markov kernel on the elements of the dual space of C_b.

- Markov kernels act on probability measures $\mu \in \mathcal{P}(\mathrm{D})$,

$$P_t\mu(A) = \int_{\mathrm{D}} P_t(x, A) \,\mathrm{d}\mu(x), \quad \text{for all } A \in \mathcal{F}.$$

Definition 2.5. Let $\{P_t\}_{t\geq 0}$ be a Markov transition kernel in D as defined before.

1. We say that μ is an *invariant measure* for the Markov process if $P_t\mu(A) = \mu(A)$ for all $A \in \mathcal{F}$ and all $t \geq 0$.

2. A Borel set $A \subseteq \mathrm{D}$ is called $(P_t, \mu)-$*invariant* if $P_t\chi_A = \chi_A$ holds $\mu-$almost everywhere, where χ_A is the indicator function for the set $A \in \mathcal{F}$.

3. An invariant measure μ is said to be *ergodic* if the unique $(P_t, \mu)-$invariant sets are D and $\varnothing$. In particular, if μ is ergodic there holds

$$\left| P_t\phi(x) - \int_{\mathrm{D}} \phi \,\mathrm{d}\mu \right| \to 0 \quad \text{as } t \to \infty$$

for all ϕ regular enough and almost all $x \in \mathrm{D}$, namely the invariant measure is attracting.

In general, if μ is the unique ergodic measure of a Markov process P_t then it is automatically ergodic. For the proof of this fact and further details about the ergodicity of Markov processes, see [32].

Next, we can proceed with a classical result about existence of invariant measures, but before we need to define an important property for measures.

Definition 2.6. We say that a collection of measures $\mathcal{M}$ on a metric space D is *tight* if for any $\varepsilon > 0$ there exists a compact subset $K_\varepsilon \subseteq \mathrm{D}$ such that

$$|\mu|(\mathrm{D} \setminus K_\varepsilon) < \varepsilon$$

for all $\mu \in \mathcal{M}$ and where $|\cdot|$ denotes the total variation.

Theorem 2.1 (Krylov–Bogolioubov)**.** *Let* D *be Polish and let* $\{P_t\}_{t\geq 0}$ *be a Markov process with the Feller property. Assume that there exists* $x_0 \in \mathrm{D}$ *such that* $P_t(x_0, \cdot)$ *is tight, then there exists at least one invariant measure* μ *on* D *for* P_t.

A constructive way to understand this statement goes as follows. Let ν be a probability measure on D, and consider the time-averaged action of P_t on ν,

$$R_t\nu = \frac{1}{t}\int_0^t P_s\nu \,\mathrm{d}s.$$

What Krylov–Bogolioubov proved is that if there exists a sequence of times $\{t_n\}_{n\in\mathbb{N}}$ and a measure μ that is a weak limit of the measures $R_{t_n}\nu$ as $n \to \infty$, then μ is an invariant measure of P_t. In addition, $R_t\nu$ will be tight if P_t is, and therefore there exists a weakly convergent subsequence. Notice that if D is compact, every collection of measures is tight and therefore the existence of an invariant measure is guaranteed.

We have seen that a sufficient condition for ergodicity is the uniqueness of an invariant measure. In relation to that we have the following classical result by Doob, [39].

Theorem 2.2 (Doob)**.** *Let $\{P_t\}_{t\geq 0}$ be a continuous Markov process and μ an invariant measure. If there exists $t_0 > 0$ such that all the transition probabilities $P_{t_0}(x, \cdot)$ are equivalent for all $x \in \mathrm{D}$, then μ is the unique invariant measure of P_t.*

Moreover we know that if $\{P_t\}_{t\geq 0}$ has the strong Feller property and is topologically irreducible, then the condition for Doob's Theorem are satisfied. However it is not trivial to prove that a Markov process has the strong Feller property, and sometimes it is more convenient to argue on a different direction. Typically, if we are hoping for exponential ergodicity a classical strategy is to look for the conditions in Harris' Theorem, see [21, 52, 67]. We elaborate on this on Chapter 5.

2.4. The mixing problem

In this section we introduce formal definitions regarding the notion of mixing of passive scalars that will be relevant for Chapter 5. In that Chapter we address the problem of ergodicity that implies some special kind of mixing, so for the sake of a better understanding we give here a brief introduction to the notion of mixing.

Consider a domain $\mathrm{D} \subset \mathbb{R}^d$, and denote by $\theta : (0, \infty) \times \mathrm{D} \to \mathbb{R}$ the passive tracer that is advected by a divergence free vector field u. Then, θ satisfies the equation

$$\partial_t \theta + u \cdot \nabla \theta = \kappa \Delta \theta,$$

where $\kappa = 0$ if there is no diffusion, i.e. transport equation, and $\kappa > 0$ in case diffusion is present, i.e. advection-diffusion equation. Without loss of generality we assume that initially $\theta(0, \cdot) = \theta^0$ has mean zero, and thus as explained in Section 2.1, $\theta(t, \cdot)$ will have mean zero for all $t > 0$ due to linearity.

Intuitively, the phenomenon of mixing can be understood as a cascading mechanism that brings information to smaller and smaller scales. A typical exemplifying situation concerns an initial configuration with black and white bits that are clearly differentiated, e.g. a 2×2 checkerboard. If we study the evolution of such configuration according to a transport or an advection-diffusion equation, there are two possible outcomes that are of interest, see Figure 2.4.

- It might happen that the black-and-white patches become more homogeneous, creating shades of grey due to the effect of the diffusion.

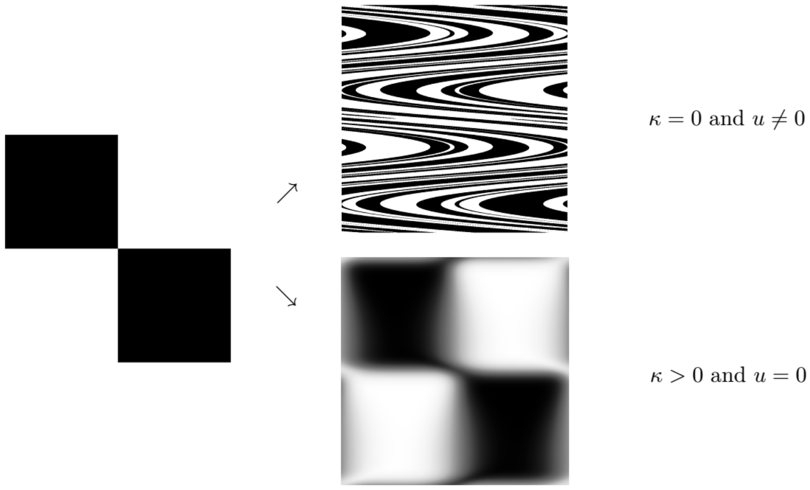

Figure 2.1.: Two types of mixing mechanisms: filamentation produced by the vector field, and homogenization produced by diffusion.

- If there is no diffusion, it might also happen that the vector field creates filaments of black-and-white areas that become thinner and thinner.

It has been a matter of interest for decades to come up with a mathematically rigorous notion of mixing that includes both possible situations, see [91]. On a first note, one can think of the variance to be a good quantity to measure mixing,

$$\mathrm{Var}(\theta) = \int_{\mathrm{D}} \left| \theta(x) - \fint \theta(z)\,\mathrm{d}z \right|^2 \mathrm{d}x.$$

However, for mean free functions there holds $\mathrm{Var}(\theta) = \|\theta\|^2_{L^2(\mathrm{D})}$, and this quantity gives no information for the $\kappa = 0$ case, since the transport equation conserves the L^2-norm in time

$$\|\theta(t)\|_{L^2(\mathrm{D})} = \|\theta(0)\|_{L^2(\mathrm{D})} \quad \text{for all } t > 0,$$

due to the divergence free property of the vector field u.

The L^2-norm or variance can be used only for the advection-diffusion equation when $\kappa > 0$. We want to find formal definitions regarding how to measure the degree of mixedness that is applicable for both the diffusive and the purely advective case.

Two decades ago, Bressan [19] introduced the so-called *geometric mixing scale*. Heuristically speaking, this notion consists on considering a magnifying glass of size $\varepsilon > 0$ and checking with it that everywhere in the domain, the average that can be seen with such magnifying glass is smaller that $\delta > 0$. More in detail we define the geometric mixing scale as follows.

Definition 2.7. Given $\varepsilon, \delta \in (0,1)$, we say that a mean zero scalar function $\theta \in L^\infty(\mathrm{D})$ is $\delta-$*mixed to scale* ε if

$$\left| \fint_{B_\varepsilon(x)\cap \mathrm{D}} \theta(z)\,\mathrm{d}z \right| \leq \delta$$

for all $x \in \mathrm{D}$.

In the original paper [19], Bressan made an important conjecture about the cost of rearrangements in $\mathrm{D} = \mathbb{T}^2$, for an initial configuration of the form

$$\theta^0 = \chi_{[0,2\pi)\times[0,\pi)} - \chi_{[0,2\pi)\times[\pi,2\pi)},$$

i.e. a 2×1 checkerboard. Bressan's conjecture says that if there is a divergence free vector field u satisfying the constraint

$$\sup_{t>0} \|\nabla u(t)\|_{L^1(\mathrm{D})} \leq 1 \tag{2.16}$$

and such that it $1/3-$mixes θ^0 to scale $\varepsilon \ll 1$ in time t, then there exists a constant $C > 0$ such that $t > C|\log \varepsilon|$. Namely, given the constrain (2.16), Bressan's mixing scale cannot decrease faster than exponentially.

The conjecture on its originally form is still open as of today, although a few years later, Crippa and DeLellis proved it to be true in [29] for vector fields that satisfy (2.16) with the L^p-norm and $p > 1$. On a recent work, Cooperman [27] proves it to be true also in the limit $p = 1$ for the special case of vector fields that are shears at each time.

The geometric mixing scale is not the only notion that is usually addressed in the literature to measure mixing. A few years after, Mathew, Mezić, Petzold, Doering and Thiffeault [38, 63] proposed negative Sobolev norms as a measure of the degree of mixedness of a passive trace. These defines the so-called *functional mixing scale* and it is the notion of mixing that we use in Chapter 5. Before proceeding with the details let us introduce some needed preliminaries.

Let $\mathrm{D} \subset \mathbb{R}^d$ a bounded domain with smooth boundary. Let $\{\phi_k\}_{k\geq 1}$ be an orthonormal basis for the mean free functions in $L^2(\mathrm{D})$ given by the Neumann–laplacian,

$$\begin{cases} -\Delta\phi_k &= \lambda_k\phi_k \quad \text{in } \mathrm{D} \\ n\cdot\nabla\phi_k &= 0 \quad\quad\;\; \text{on } \partial\mathrm{D}, \end{cases}$$

with $\|\phi_k\|_{L^2} = 1$ for all integers $k \geq 1$. Since we are removing the zero-th mode because of the mean free condition, notice that $\lambda_k > 0$ for all $k \geq 1$ and moreover there holds

$$\lambda_k \leq \lambda_{k+1} \quad \text{for all integers } k \geq 1.$$

With this notion we define the *homogeneous Sobolev space of order* m, as the set of measurable and mean zero functions $f : \mathrm{D} \to \mathbb{R}$ such that

$$\|f\|^2_{\dot{H}^m(\mathrm{D})} = \sum_{k \geq 1} \lambda_k^m \langle \phi_k, f \rangle^2_{L^2(\mathrm{D})} < \infty.$$

Notice that $\{\phi_k\}_{k\geq 1}$ also forms an orthogonal basis in $\dot{H}^m(\mathrm{D})$ since $\lambda_k > 0$ for all $k \geq 1$ therefore $\lambda_k \langle \phi_k, f \rangle = 0$ for all $k \geq 1$ implies $f = 0$. This definition is motivated by the fact that in this way we can write

$$\|\phi_k\|^2_{\dot{H}^m(\mathrm{D})} = \|\nabla^m \phi_k\|^2_{L^2(\mathrm{D})} = \lambda_k^m \|\phi_k\|^2_{L^2(\mathrm{D})} = \lambda_k^m$$

for any integer $k \geq 1$ and $m \in \mathbb{R}$.

In addition, notice that the negative homogeneous Sobolev norms are a good measure of the degree of mixedness of a passive scalar. Each mode of the Sobolev norm of order $m = -s < 0$ scales as

$$\frac{\langle \phi_k, f \rangle^2}{k^{2s/d}},$$

and therefore a convergence to zero of such norm can be a consequence of two phenomena.

- *Homogenization.* It could be that each mode $\langle \phi_k, f \rangle$ individually goes to zero, which means that all the energy is dissipating and thus it can only happen if there is diffusion $\kappa > 0$.
- *Filamentation.* The $\dot{H}^{-s}$ norm can decay as well if all the energy goes to the larger–in–k modes $\langle \phi_k, f \rangle$ since the higher $k-$modes are divided by a larger factor $k^{2s/d}$. In such case the total energy, namely the L^2 norm, remains conserved and thus this occurs only if there is no diffusion $\kappa = 0$.

With that in mind we define the functional mixing scale, which is used as a measure of mixing in Chapter 5.

Definition 2.8. We say that a vector field u *mixes* $\theta^0 \in (L^\infty \cap H^s)(\mathrm{D})$ mean zero if there exists $s > 0$ such that the solution θ to the transport equation starting from θ^0 satisfies

$$\|\theta(t)\|_{\dot{H}^{-s}(\mathrm{D})} \to 0 \quad \text{as } t \to \infty.$$

In particular we say that there is *exponential mixing* if there exists $\gamma > 0$ such that

$$\|\theta(t)\|_{\dot{H}^{-s}(\mathrm{D})} \lesssim \|\theta^0\|_{H^s(\mathrm{D})} e^{-\gamma t}$$

for all $t > 0$.

The particular choice of $s > 0$ for the $\dot{H}^{-s}$−norm is not so relevant. Originally the most frequent norm was the $\dot{H}^{-1/2}$−norm [40, 63], since it is equivalent to some previous definition of mix-norm that is more related to Bressan's geometric mixing scale. More recently, the most prevailing way to define the mixing scales is by means of the $\dot{H}^{-1}$−norm, [1, 62, 68]. Notice that the $\dot{H}^{-s}$−norm scales like $[\text{Length}]^s$, and thus if $s = 1$ the norm potentially has a physical interpretation as the *length of the filamentations*. There are some other relevant choices of functional mixing scales, e.g. optimal transport distances [82], that can be related to some negative Sobolev norms but that we will not discuss them here.

3. Stability results for distributional solutions with rough vector fields

This chapter is based on the article [72], which is a joint work with André Schlichting and Christian Seis. Large parts of it are reproduced verbatim.

Chapter summary

This work contains two main contributions. First, it provides optimal stability estimates for advection-diffusion equations in a setting in which the velocity field is Sobolev regular in the spatial variable. This estimate is formulated with the help of Kantorovich–Rubinstein distances with logarithmic cost functions. Second, the stability estimates are extended to the advection-diffusion equations with velocity fields whose gradients are singular integrals of L^1 functions, entailing a new well-posedness result.

3.1. Introduction

In this section we deal with stability properties for solution to the advection-diffusion equation. Consider a passive scalar θ in $\mathbb{R}^d$ driven by some vector field $u : (0,T)\times\mathbb{R}^d \to \mathbb{R}^d$ and in presence of diffusion $\kappa > 0$. Then the dynamics of the passive scalar are described by the Cauchy problem,

$$\begin{cases} \partial_t\theta + \nabla\cdot(u\theta) &= \kappa\Delta\theta \qquad \text{in } (0,T)\times\mathbb{R}^d, \\ \theta(0,\cdot) &= \theta^0 \qquad \text{in } \mathbb{R}^d, \end{cases} \tag{3.1}$$

where θ^0 represents the initial configuration.

The mathematical theory in the setting of smooth vector fields is contained in the general classical theory for parabolic equations; see, for instance, Ladyzhenskaya et al. [58]. Thanks to the linearity of the equations, well-posedness is then based on simple a priori estimates.

However, there are some important examples, for instance, in the areas of fluid dynamics or kinetic theories, in which the advecting velocity fields and the observed quantities are rather rough. Thus, the mathematical theory for the advection-diffusion equation (3.1) falls out of the setting covered by classical theories, and it has been discovered only recently that there are situations in which integrable distributional solutions cease to be unique [22, 70, 69]. For more details about uniqueness of solutions in different regularity settings see Section 2.1

A customary proof of uniqueness is based on energy estimates. We are aware of two approaches to energy estimates that deal with different weak hypotheses on regularity and integrability of vector fields and solutions, both embarking from the following (at first formal) estimates

$$\frac{1}{(q-1)q}\frac{d}{dt}\int_{\mathbb{R}^d}|\theta|^q\,\mathrm{d}x+\kappa\int_{\mathbb{R}^d}|\theta|^{q-2}|\nabla\theta|^2\,\mathrm{d}x\le\begin{cases}\int_{\mathbb{R}^d}|\theta|^{q-1}|\nabla\theta||u|\,\mathrm{d}x,\\[2ex]\int_{\mathbb{R}^d}|\theta|^q(\nabla\cdot u)^-\,\mathrm{d}x,\end{cases}\tag{3.2}$$

for $q>1$. Here, the superscript minus sign labels the negative part of the divergence. The limit $q\to 1$ leads to an a priori estimate in terms of the entropy and is discussed in Remark 3.1 below.

The first approach, which is based on the first estimate, applies to velocity fields in the integrability class $L^r((0,T);L^p(\mathbb{R}^d))$ provided that r and p satisfy the so-called Ladyzhenskaya–Prodi–Serrin condition

$$\frac{2}{r}+\frac{d}{p}\le 1.$$

Here, the task is to bound the integral on the right-hand side in terms of those on the left-hand side, which can be achieved with standard Hölder and Sobolev inequalities. Apparently, the parabolic structure is of fundamental importance in this approach and the method ceases to hold in the non-diffusive setting $\kappa=0$. Since we are particularly interested in estimates that hold uniformly for positive but arbitrary small diffusivity parameter κ, we will not further elaborate on it here. We refer to [11] for a simple proof in the $q=2$ setting and a discussion on optimality.

The second approach is particularly important in models in which the fluid is at most weakly compressible in the sense that

$$(\nabla\cdot u)^-\in L^1\big((0,T);L^\infty(\mathbb{R}^d)\big).\tag{3.3}$$

In this case, an application of a Gronwall argument implies the estimate

$$\|\theta\|_{L^\infty(L^q)} + c_{\kappa,q}\|\nabla|\theta|^{\frac{q}{2}}\|_{L^2(L^2)}^{\frac{2}{q}} \lesssim \Lambda^{1-\frac{1}{q}}\|\theta^0\|_{L^q} \tag{3.4}$$

where $c_{\kappa,q}^q = \kappa(q-1)$ and $\log\Lambda = \|(\nabla\cdot u)^-\|_{L^1(L^\infty)}$. In order to make this approach work rigorously, the validity of the chain rule has to be confirmed to establish the energy estimate (3.2). This leads us to the concept of *renormalized solutions*, which were originally introduced by DiPerna and Lions in [37]: An integrable function θ is called a renormalized solution to the advection-diffusion equation (3.1) if it satisfies

$$\partial_t\beta(\theta) + \nabla\cdot(u\beta(\theta)) = (\beta(\theta) - \theta\beta'(\theta))\nabla\cdot u + \kappa\Delta\beta(\theta) - \kappa\beta''(\theta)|\nabla\theta|^2 \tag{3.5}$$

in the distributional sense, for any bounded C^2 function $\beta : \mathbb{R} \to \mathbb{R}$ whose derivatives are bounded and vanish at zero. Renormalized solutions are easily proved to be unique, and DiPerna and Lions' theory shows that distributional solutions in $L^\infty((0,T); L^q(\mathbb{R}^d))$ are renormalized if the advecting velocity field is Sobolev regular in the spatial coordinate, namely $u \in L^1((0,T); W^{1,p}(\mathbb{R}^d))$, and p and q have to be Hölder conjugates (or larger), $1/p + 1/q \leq 1$. In what follows, we will occasionally refer to this setting as the *DiPerna–Lions setting*.

The DiPerna–Lions theory was further extended by Ambrosio [3] to vector-fields of BV-regularity. However, there are certain situations in which a direct verification of the renormalization property (3.5) seems to fail. One regards well-posedness results for vector-fields, whose gradient is a singular integral of an L^1 function [31], which is of interest in certain problems in fluid dynamics. We revisit this setting later in this paper.

It is certainly surprising that the regularizing effect of diffusion does not rule out non-uniqueness and that the DiPerna–Lions setting is *both* optimal for the advection equation ($\kappa = 0$ in (3.1)) as well as the advection-diffusion equation (3.1), at least up to a dimension-dependent gap.

DiPerna and Lions' theory is extremely powerful and finds numerous applications to various types of advection and kinetic equations. As a by-product, it provides qualitative stability statements for the linear equation (3.1). For instance, considering two different solutions with two different advection fields,

$$\partial_t\theta_\varepsilon + \nabla\cdot(u_\varepsilon\theta_\varepsilon) = \kappa\Delta\theta_\varepsilon, \qquad \partial_t\theta + \nabla\cdot(u\theta) = \kappa\Delta\theta,$$

we know that $\theta_\varepsilon \to \theta$ when $\varepsilon \to 0$ provided $u_\varepsilon \to u$ in some suitable norms. Similarly, for vanishing diffusivities, $\kappa \to 0$, solutions of the advection-diffusion equation (3.1) converge to the solutions of the transport equation

$$\partial_t \theta + \nabla \cdot (u\theta) = 0. \tag{3.6}$$

By nature, DiPerna and Lions' theory cannot provide rates in these qualitative stability statements. Besides establishing well-posedness, in particular uniqueness, those are interesting from the point of view of an error analysis for numerical approximations. But also for modeling purposes, quantitative results are crucial, for instance, with regard to the zero-diffusivity limit $\kappa \to 0$.

The purpose of this chapter is to derive *stability estimates* for the advection-diffusion equation (3.1) both in the DiPerna–Lions setting and the slightly more singular setting from [31].

The works [83, 84] provide a new quantitative approach to the advection equation (5.1) in the DiPerna–Lions setting. The approach not only rediscovers most of the results from the original paper [37] but also offers sharp stability estimates that extend to situations inaccessible by the traditional renormalization approach. A typical estimate in this context compares two distributional solutions $\theta_1, \theta_2 \in L^\infty((0,T); L^q(\mathbb{R}^d))$ of the Cauchy problem (5.1) corresponding to two different weakly compressible velocity fields $u_1, u_2 \in L^1((0,T); L^p(\mathbb{R}^d))$, respectively, where $1/p + 1/q = 1$. Then it holds,

$$\sup_{t\in(0,T)} \mathcal{D}_\delta(\theta_1(t,\cdot), \theta_2(t,\cdot)) \lesssim \|\nabla u_1\|_{L^1(L^p)} \big(\|\theta_1\|_{L^\infty(L^q)} + \|\theta_2\|_{L^\infty(L^q)}\big) + 1, \tag{3.7}$$

provided that $u_1 \in L^1((0,T); W^{1,p}(\mathbb{R}^d))$ and where $\delta = \|u_1 - u_2\|_{L^1(L^p)}$ is the distance of the velocity fields. The quantity $\mathcal{D}_\delta(\cdot,\cdot)$ on the left-hand side is a Kantorovich–Rubinstein distance associated to a logarithmically increasing cost, originally arising in optimal transportation theory and defined as

$$\mathcal{D}_\delta(\mu, \nu) = \inf_{\pi\in\Pi(\mu,\nu)} \iint_{\mathbb{R}^d\times\mathbb{R}^d} \log\Big(\frac{|x-y|}{\delta} + 1\Big)\, d\pi(x,y). \tag{3.8}$$

Here, μ, ν are finite measures on $\mathbb{R}^d$ such that $\mu[\mathbb{R}^d] = \nu[\mathbb{R}^d]$ and $\Pi(\mu,\nu)$ is the set of couplings, i.e., measures on $\mathbb{R}^d \times \mathbb{R}^d$ such that $\pi[A \times \mathbb{R}^d] = \mu[A]$ and $\pi[\mathbb{R}^d \times A] = \nu[A]$ for all measurable $A \subset \mathbb{R}^d$. The parameter $\delta > 0$ here plays a crucial role because it can be understood as the rate of convergence between the two densities θ_1, θ_2 in terms of some parameter (the $L^1((0,T); L^p(\mathbb{R}^d))$ distance between u_1 and u_2 in this case). See Section 2.2 for more details about this topic and related optimal transport distances.

The first version of the stability estimate (3.8) was introduced in [18], and we shall comment briefly on its origin. The fact that there are logarithmic distances appearing is not really surprising since they are already present at the level of the flow. Consider the ODE associated to the transport equation without diffusion, that is the *Lagrangian setting*, as the equation for the flow $\phi_t : \mathbb{R}^d \to \mathbb{R}^d$ with $t \in (0,T)$,

$$\begin{cases} \partial_t \phi_t = u_t \circ \phi_t, \\ \phi_0 = \mathrm{id}\,. \end{cases} \tag{3.9}$$

Then, for two solutions of the flow equation (3.9) there is also the following logarithmic stability estimate that can be verified straightforwardly,

$$\log\left(\frac{|\phi_t^1(x) - \phi_t^2(x)|}{\delta} + 1\right) \lesssim \|\nabla u_1\|_{L^1(L^\infty)} + 1, \tag{3.10}$$

provided that $\delta = \|u_1 - u_2\|_{L^1(L^\infty)}$ and that u_1 is Lipschitz. Analogous results in the DiPerna–Lions setting for the non-diffusive case have been proved by Crippa and De Lellis [29] if $p > 1$ by replacing the uniform bounds in the space variable by suitable integral averages. Unfortunately, due to technical limitations, it is currently unclear how to extend these *optimal* stability estimates to the case $p = 1$. There are, however, non-optimal extensions of the Crippa–De Lellis theory to the case $p = 1$ by Jabin [55] and to lower regular vector fields, namely vector fields whose gradient is given by a singular integral of an L^1 function, by Bouchut and Crippa [15].

The ordinary differential equation of the flow (3.9) can be related to the transport equation (5.1) through the method of characteristics,

$$\theta(t,\cdot) = (\phi_t)_\# \theta^0,$$

and therefore the existence of an analogy between (3.7) and (3.10) is not unexpected.

The stability estimates for the transport equation [83, 84] turned out to be quite flexible. There are actually many applications for which the optimal estimate plays a crucial role, for example in coarsening and mixing problems [18, 75, 82, 85] or when studying error estimates for numerical approximations [79, 80]. Moreover, stability estimates are successfully extended to certain settings, in which renormalization in the sense of DiPerna and Lions could not be established directly [25, 31]. As a consequence, new well-posedness results are proven. In the stochastic setting, (sub-optimal) stability estimates were derived in [44, 61].

In this chapter we intend to generalize these optimal stability estimates to transport equations with diffusivity, i.e., $\kappa > 0$. While the regularizing effect of the diffusion might on a qualitative level indicate that estimates holding for the transport equation (5.1) carry over to the advection-diffusion equation (3.1), the adaptation of the mathematical proofs is not straightforward. As it turns out, our analysis is limited to advection-diffusion equations with constant diffusivities, while qualitative results are available for more general diffusions [45, 59, 60]. Nonetheless, these estimates are an important contribution to the existing theory. For instance, optimal bounds on convergence rates of numerical approximations become accessible for the first time in the DiPerna–Lions setting, since it is addressed in [71] and presented here in Chapter 4. Moreover, the new bounds are potentially applicable in the study of mixing problems in fluid dynamics, see also [82, 85] for related work. In addition, our estimate extends the existing well-posedness theory for (3.1) to fluids with an L^1 vorticity, $\nabla \times u \in L^1$, which are of relevance, for instance, in the study of the $2D$ Euler and Navier–Stokes equations [24, 31, 73].

We finally mention that quantitative estimates on advection-diffusion equation were also obtained in the recent works [20, 85]. These, however, focus on quantifying the vanishing diffusivity limit with applications to mixing.

The chapter is organized as follows: In the next section, Section 3.2, we state the precise definitions and present and discuss our main results. In Section 3.3, we present the proof of our general stability estimate in Theorem 3.1. The final Section 3.4 contains a uniqueness result for vector fields with L^1 vorticites, Theorem 3.2.

3.2. Main results

Our first main result of this chapter concerns a stability estimate for the advection-diffusion equation (3.1) using the optimal transportation distance (3.8). For this, we are considering precisely the DiPerna–Lions setting [37] that we introduced before, that is, for the velocity field, we impose Sobolev regularity in the spatial variable,

$$u \in L^1((0,T); W^{1,p}(\mathbb{R}^d)) \qquad \text{for some } p \in (1, \infty], \tag{3.11}$$

while for the initial datum, we demand some integrability,

$$\theta^0 \in (L^1 \cap L^q)(\mathbb{R}^d) \qquad \text{with } q > 1 \text{ such that } \frac{1}{p} + \frac{1}{q} \leq 1. \tag{3.12}$$

Working in the full space requires to suppose some additional decay properties, for instance, in order to ensure that the logarithmic Kantorovich–Rubinstein norms $D_\delta(\theta^0)$ are finite for *any finite* δ. We achieve this by additionally assuming that θ^0 has finite first moments,

$$\int_{\mathbb{R}^d} |x| \, |\theta^0(x)| \, \mathrm{d}x < \infty. \tag{3.13}$$

In this setting, it can be established that (smooth) solutions have L^1 temporal-spatial gradients as will be outlined in Remark 3.1 below. On bounded domains, the latter is always true as a consequence of the a priori estimates in (3.4).

Let us now present the precise statement.

Theorem 3.1. *Let u_1 and u_2 be two vector fields satisfying* (3.11)*, and let θ_1^0 and θ_2^0 be two initial data such that* (3.12) *holds. Let κ_1 and κ_2 be two positive constants. Then, for any two distributional solutions $\theta_1, \theta_2 \in L^\infty(L^1 \cap L^q)$ to the advection-diffusion equation* (3.1) *corresponding to κ_1, u_1, θ_1^0 and κ_2, u_2, θ_2^0, respectively, which satisfy $\theta_1, \theta_2 \in L^1(W^{1,1})$, the following stability estimate holds*

$$\sup_{0\le t\le T} \mathcal{D}_\delta(\theta_1,\theta_2)(t) \lesssim 1 + \mathcal{D}_\delta(\theta_1^0,\theta_2^0) + \frac{\|u_1-u_2\|_{L^1(L^p)} + |\kappa_1-\kappa_2|\,\|\nabla\theta_2\|_{L^1(L^1)}}{\delta}. \quad (3.14)$$

It is not difficult to see that the stability estimate implies uniqueness. Indeed, suppose θ_1 and θ_2 are two solutions to the advection-diffusion equation with the same velocity, initial data, and diffusivity constant. In that case, the right-hand side becomes independent of δ and letting $\delta \to 0$; the left-hand side would blow up except if both solutions are identical. The argument could be made rigorous, for instance, by a straightforward application of Lemma 2.1 below, and we will elaborate on this principle in the proof of Theorem 3.2. We thus recover the uniqueness results for distributional solutions from the original paper by DiPerna and Lions [37] in a new quantitative way.

The implicit constant in the stability estimate (3.14) depends on the $L^\infty(L^1)$ and $L^\infty(L^q)$ norms of both solutions, but only on the $L^1(L^p)$ norm of one of the velocity gradients and the $L^1(L^1)$ norm of one of the solutions gradients. Consequently, it would be enough to assume such regularity for one of the vector fields and one of the solutions, and we would get an estimate on the distance between the unique solution and a non-unique approximant. Furthermore, our analysis applies also to the situation where one of the diffusivity constants depends on x, in which case the modulus of the difference of the diffusivity constants needs to be replaced by $\|\kappa_1 - \kappa_2(x)\|_{L^\infty}$.

Because Kantorovich–Rubinstein distances metrize weak convergence, cf. Theorem 7.12 in [94], the result of Theorem 3.1,

$$\sup_{0\le t\le T} \mathcal{D}_\delta(\theta(t,\cdot),\theta_n(t,\cdot)) \lesssim \mathcal{D}_\delta(\theta^0,\theta_n^0) + 1 + \frac{\|u-u_n\|_{L^1(L^p)} + |\kappa-\kappa_n|}{\delta},$$

cf. (3.14), can be considered as an estimate on the rates of weak convergence for the advection-diffusion equation for three types of approximations: The convergence of solution sequences corresponding to converging sequences of initial data $\theta_n^0 \to \theta^0$, velocity fields $u_n \to u$ and diffusivity constants $\kappa_n \to \kappa$. Here, an interesting feature is that the rate of convergence is already incorporated into the distance function, and it is given by the smallest $\delta = \delta_n$ for which the right-hand side is finite, i.e.,

$$\mathcal{D}_{\delta_n}(\theta^0, \theta_n^0) \lesssim 1, \qquad \|u - u_n\|_{L^1(L^p)} \lesssim \delta_n, \qquad |\kappa - \kappa_n| \lesssim \delta_n. \tag{3.15}$$

A setting in which all three of the above errors occur is the numerical approximation of (3.1), which we study in a parallel work in [71]. Indeed, a finite volume scheme introduces three discretization errors on the level of the initial datum, the velocity field and due to numerical diffusion also on the level of the diffusivity constant. The latter observation motivates the investigation of the stability in the limit $\kappa_n \to \kappa > 0$. Moreover, for a finite volume discretization of a Sobolev vector field u on a mesh of size $1/n$, the rate $\delta_n = 1/n$ in (3.15) is shown to be optimal [71].

The numerical application also highlights the need to work with weak norms, since if θ_n^0 denotes a discretization of a merely integrable initial datum θ^0 on a mesh of size $1/n$, the approximation is converging weakly with rate $1/n$, see, e.g., [79, 80] in the case of a finite volume approximation, but there is no rate of convergence in any strong Lebesgue norm. Therefore, in order to estimate the approximation error for numerical schemes for the advection-diffusion equation in the DiPerna–Lions setting, it is necessary to choose distances which metrize weak convergence such a negative Sobolev norms or Kantorovich–Rubinstein distances. For us, this is the main motivation for the derivation of Theorem 3.1.

Stability estimates for this type of perturbations are derived in the Lagrangian setting, which correspond here to stochastic differential equations, in [97], based on the ODE stability obtained by Crippa and De Lellis in [29]. In the Eulerian setting, similar results are derived by Li and Luo [61], building on the stability theory for advection equations developed in [83, 84]. In all these works, the authors obtain only the suboptimal $\sqrt{|\kappa_n - \kappa|}$ error by treating the diffusion as a perturbation to the advection.

In the present chapter, our objective is to improve the rate of weak convergence from $\sqrt{|\kappa_n - \kappa|}$ to $|\kappa_n - \kappa|$, in order to estimate optimally the numerical error induced by finite volume schemes for the advection-diffusion equation, cf. [71]. The improvement from "exponent 1/2 convergence" to "exponent 1 convergence" is well-known by numerical analysts who study numerical schemes for advection-diffusion equations in a more regular setting, see e.g. [43, Chapter 4]. It is a consequence of the smoothing effect of diffusion, which is best understood if we neglect the advection for a moment: In this case, the zero-diffusivity limit corresponds to the convergence of smooth solutions to its *rough* initial configuration, while in the case of positive diffusivity, *smooth* solutions are approximated.

Besides the indication of optimality in the numerical application, the following simple example vaguely indicates that our findings are optimal. For simplicity, we consider the one-dimensional case only and a Dirac function as initial datum. Moreover, we trade the logarithmic Kantorovich–Rubinstein distance for the Wasserstein distance W_1, see (2.8) below, because computations here can be made more explicit.

Example 3.1. Suppose that θ_1 and θ_2 denote the one-dimensional heat kernels corresponding to diffusivities $\kappa_1 > \kappa_2 > 0$. Then it holds

$$\frac{t|\kappa_1 - \kappa_2|}{\sqrt{\kappa_1} + \sqrt{\kappa_2}} \lesssim W_1(\theta_1(t,\cdot), \theta_2(t,\cdot)).$$

Indeed, since the Wasserstein distance is the supremum over all Lipschitz functions $\psi(x)$ with Lipschitz constant bounded by 1, cf. (2.8) below, we can consider the specific function $\psi(x) = |x|$ to produce an explicit lower bound. Then, by means of the change of variables $x = 2\sqrt{t\kappa_i}y$ for $i = 1, 2$ we obtain

$$\begin{aligned} W_1(\theta_1(t,\cdot), \theta_2(t,\cdot)) &\geq \int_{\mathbb{R}^d} (\theta_1(t,x) - \theta_2(t,x))\psi(x)\ \mathrm{d}x \\ &= \frac{1}{(4\pi\kappa_1 t)^{d/2}} \int_{\mathbb{R}^d} |x| e^{-\frac{|x|^2}{4\kappa_1 t}}\ \mathrm{d}x - \frac{1}{(4\pi\kappa_2 t)^{d/2}} \int_{\mathbb{R}^d} |x| e^{-\frac{|x|^2}{4\kappa_2 t}}\ \mathrm{d}x \\ &= \frac{2\sqrt{t}}{\pi^{d/2}} (\sqrt{\kappa_1} - \sqrt{\kappa_2}) \int_{\mathbb{R}^d} |y| e^{-|y|^2}\ \mathrm{d}y \gtrsim \frac{\sqrt{t}}{\sqrt{\kappa_1} + \sqrt{\kappa_2}} (\kappa_1 - \kappa_2). \end{aligned}$$

Of course, despite the fact that the Kantorovich–Rubinstein distance $\mathcal{D}_\delta$ and the Wasserstein distance W_1 both metrize weak convergence, we are aware of the fact that they are not equivalent and thus the actual rate of convergence might be different. We have the one-sided estimate $\mathcal{D}_\delta \leq W_1/\delta$ which follows from linearizing the logarithm. However, the opposite estimate, that would be desirable here, does not hold true. Moreover, also by direct calculations, it is not clear to us how to extend the example to the $\mathcal{D}_\delta$ distance.

Let us briefly comment on the vanishing viscosity limit $\kappa_n \to 0$. While strong convergence can be easily established, simple examples show that without imposing additional regularity on the initial datum, there cannot be a convergence rate [84]. To obtain rates of strong convergence very mild (e.g. logarithmic Sobolev) regularity assumption on the initial datum are sufficient [20, 14, 66]. In comparison, the optimal rate of weak convergence is of the order $\sqrt{\kappa}$, as established in [83]. This limit also plays a role in the analysis of numerical approximations for the purely advective equation, $\kappa = 0$, see [79, 80].

We finally remark on our hypotheses in Theorem 3.1. For initial data in (3.12) and velocity fields in (3.11), the existence of distributional solutions in $L^\infty(L^1 \cap L^q)$ is easily established with the help of the a priori estimates in (3.4) via smooth approximation, *provided* that the velocity field satisfies the weak compressibility condition in (3.3). In the following remark, we comment on the gradient condition.

Remark 3.1. The regularity assumption $\nabla\theta \in L^1(L^1)$ that is assumed in Theorem 3.1 is satisfied for finite entropy solutions as long as the velocity verifies the weak compressibility condition in (3.3). Indeed, if θ is a nonnegative solution with

$$\int_{\mathbb{R}^d} \theta(t,x)\log\theta(t,x)\,\mathrm{d}x \in \mathbb{R} \qquad \text{for all } t\in[0,T], \tag{3.16}$$

a standard computation reveals that

$$\int_{\mathbb{R}^d} \theta(t)\log\theta(t)\,\mathrm{d}x + \kappa\int_0^t\int_{\mathbb{R}^d} \frac{|\nabla\theta|^2}{\theta}\,\mathrm{d}x\,\mathrm{d}t \le \int_{\mathbb{R}^d}\theta^0\log\theta^0\,\mathrm{d}x + \|(\nabla\cdot u)^-\|_{L^1(L^\infty)}\|\theta\|_{L^\infty(L^1)},$$

and thus, the Fisher information $\|\theta^{-1}|\nabla\theta|^2\|_{L^1}$ is integrable in time. Moreover, by Hölder's inequality, we obtain

$$\int_0^t \|\nabla\theta\|_{L^1}^2\,\mathrm{d}s \le \int_0^t \|\theta\|_{L^1}\|\theta^{-1}|\nabla\theta|^2\|_{L^1}\,\mathrm{d}s \le \|\theta\|_{L^\infty(L^1)}\int_0^t \|\theta^{-1}|\nabla\theta|^2\|_{L^1}\,\mathrm{d}s,$$

which is finite by (3.4), (3.16), and the above estimate on the entropy. We easily deduce that $\nabla\theta \in L^1(L^1)$ on finite time intervals, such that the last term in the right hand side of (3.14) can be estimated by

$$\frac{|\kappa_1-\kappa_2|\;\|\nabla\theta_2\|_{L^1(L^1)}}{\delta} \lesssim \frac{|\kappa_1-\kappa_2|}{\delta\,\kappa_2}.$$

It remains to understand that in the setting (3.3), (3.11), (3.12), (3.13), solutions do indeed have finite entropy for finite times (3.16). An upper bound is provided by the elementary estimate $r\log r \lesssim r + r^q$ for any $r>0$ and the integrability assumptions on θ. For the lower bound, we first notice that the moment bound in (3.13) is propagated in time. In fact, since $\theta^0\in L^1$ by assumption, the homogeneous weight $|x|$ in (3.13) can be replaced by the smoother $\sqrt{1+|x|^2}$, and we have the estimate

$$\int_{\mathbb{R}^d}\sqrt{1+|x|^2}\theta(t,x)\,\mathrm{d}x \lesssim \int_{\mathbb{R}^d}\sqrt{1+|x|^2}\theta^0(x)\,\mathrm{d}x + \|u\|_{L^1(L^p)}\|\theta\|_{L^\infty(L^q)} + \kappa\|\theta\|_{L^1}.$$

Now, since $r\log r \gtrsim -r^\alpha$ for any $\alpha\in(0,1)$ and any $r>0$, we conclude that the lower bound

$$\int_{\mathbb{R}^d} \theta \log \theta \, \mathrm{d}x \gtrsim - \int_{\mathbb{R}^d} \theta^\alpha \, \mathrm{d}x \gtrsim - \left(\int_{\mathbb{R}^d} \frac{1}{\sqrt{1+|x|^2}^{\frac{\alpha}{\alpha-1}}} \, \mathrm{d}x \right)^{1-\alpha} \left(\int_{\mathbb{R}^d} \sqrt{1+|x|^2} \, \theta \, \mathrm{d}x \right)^\alpha$$

is finite as long as $\alpha > \frac{d}{d+1}$.

We complete the discussion of Theorem 3.1 with a comment on the integrability restriction on the velocity gradient in (3.11). A crucial step in the derivation of the stability estimate is controlling the advection term via an argument that was introduced by Crippa and De Lellis in [29]. The argument makes use of the Hardy–Littlewood maximal function and exploits its continuity in L^p spaces for $p > 1$. The restriction in (3.11) relies precisely on this limitation. See Section 3.3 for details.

In our second theorem, that we shall motivate in the following, we use a suitable extension of the Crippa–De Lellis method to vector fields whose gradient is given by a singular integral of an L^1 function. A typical example of such a vector field is the velocity field that is obtained from an L^1 vorticity with the help of the Biot–Savart law.

The precise setting is as follows. We assume that the velocity components $u_1, \dots, u_d$ have kernel representations,

$$u_i = k_i * \omega_i, \tag{3.17}$$

for any $i \in \{1, \dots, d\}$, where the generalized vorticity components are merely integrable,

$$\omega_i \in L^1((0,T); L^1(\mathbb{R}^d)), \tag{3.18}$$

and where the kernels k_i are such that any of its derivatives $\partial_j k_i$ is a *singular kernel*. More precisely, we suppose that

(k$_1$) $k \in \mathcal{S}'(\mathbb{R}^d)$, where $\mathcal{S}'(\mathbb{R}^d)$ is the dual of the Schwartz space;

(k$_2$) $k|_{\mathbb{R}^d \setminus \{0\}} \in C^2(\mathbb{R}^d \setminus \{0\})$;

(k$_3$) for $\alpha \in \mathbb{N}_0^d$ with $|\alpha| \le 2$ it holds

$$|D^\alpha k(x)| \lesssim \frac{1}{|x|^{d-1+|\alpha|}}, \qquad \forall x \neq 0;$$

(k$_4$) it holds

$$\left| \int_{R_1 < |x| < R_2} \nabla k(x) \, \mathrm{d}x \right| \lesssim 1 \qquad \text{for every } 0 < R_1 < R_2 < \infty.$$

Under these conditions, the velocity field only lives in a *weak* Lebesgue space globally in $\mathbb{R}^d$, see Lemma 3.2 below. For mathematical convenience, we will enforce the slightly stronger condition

$$u \in L^{p,\infty}((0,T) \times \mathbb{R}^d), \tag{3.19}$$

for some $p > 1$. For the above mentioned applications in fluid dynamics, such a condition is always satisfied. We recall that weak Lebesgue spaces $L^{p,\infty}$ can be defined on a measure space (X, μ) via the quasi-norm

$$\|f\|_{L^{p,\infty}(\mu)} = \sup_{\lambda>0} \left(\lambda^p \mu(\{x \in X : |f(x)| > \lambda\})\right)^{1/p}, \tag{3.20}$$

for any measurable $f : X \to \mathbb{R}$. In the case $p = \infty$, we adopt the convention $L^{\infty,\infty} = L^\infty$. Notice that $\|\cdot\|_{L^{p,\infty}}$ is not a norm since it does not verify the triangle inequality. Also recall that there is an embedding $L^p \subset L^{p,\infty}$ with $\|f\|_{L^{p,\infty}} \leq \|f\|_{L^p}$ for every $f \in L^p$.

As our goal is to derive a full well-posedness theorem, and not only a uniqueness result, we shall, in addition, assume that the velocity field satisfies the weak compressibility condition (3.3).

Finally, we work with with initial datum that are integrable and bounded,

$$\theta^0 \in (L^1 \cap L^\infty)(\mathbb{R}^d), \tag{3.21}$$

and that have finite first moments, i.e., (3.13) holds.

With this list of assumptions, we can state our second main result.

Theorem 3.2. *Let u be a velocity field satisfying* (3.17)*,* (3.18)*,* (3.19)*, and* (3.3)*, and let θ^0 be an initial datum verifying* (3.21) *and* (3.13)*. Then the Cauchy problem* (3.1) *has a unique distributional solution $\theta(t,x)$ in the class $L^\infty((0,T);(L^1 \cap L^\infty)(\mathbb{R}^d))$ with $\nabla\theta \in L^1((0,T) \times \mathbb{R}^d)$.*

A corresponding result on uniqueness for the transport equation (5.1) was previously derived in [31], which in turn builds up on the Lagrangian setting considered in [15]. Here, we develop an analogous theory for the diffusive case $\kappa > 0$. The result is, of course, not unexpected since the diffusive equation is usually considered to produce even smoother solutions. However, we are currently not aware of any techniques, apart from those developed here, in which such a result can be established. Moreover, we are presently unable to produce results in more general settings, for instance, for non-constant diffusivities.

The new uniqueness result in Theorem 3.2 has potential applications in the study of the inviscid limit for the two-dimensional Navier–Stokes equations with rough forcing.

We believe that thanks to our present contribution, the recent results in [73, 24] can be extended to such an interesting setting.

Finally, in Theorem 3.2 we add a proof of existence that is based on classical techniques. Assumptions on the advection field itself and the weak compressibility assumption are only used to establish the existence part.

3.3. Optimal stability results in the DiPerna–Lions setting

Next in order, we will introduce a first tool that is essential in the proof of Theorem 3.1. It is based on the Hardy–Littlewood maximal function operator M, which is a central tool from the Calderón–Zygmund theory defined for any measurable functions $f:\mathbb{R}^d\to\mathbb{R}$ by

$$Mf(x)=\sup_{R>0}\frac{1}{|B_R(x)|}\int_{B_R(x)}|f(z)|\,\mathrm{d}z. \tag{3.22}$$

The operator is continuous from L^p to L^p if $p\in(1,\infty]$, hence $\|Mf\|_{L^p}\lesssim\|f\|_{L^p}$, see [87] for details. Moreover with the maximal function operator one can find bounds for the difference quotients by using Morrey-type inequalities, namely, for almost any $x,y\in\mathbb{R}^d$ it holds

$$\frac{|f(x)-f(y)|}{|x-y|}\lesssim M(\nabla f)(x)+M(\nabla f)(y). \tag{3.23}$$

A proof of the continuity of the Morrey-type estimate can be found in [42]. The estimate in Lemma 3.1 below resembles the one in [83, Lemma 3], but here we adapt the setting and the notation to make it more useful for our purpose.

Lemma 3.1. *Let $p\in(1,\infty]$ and $q\in[1,\infty)$ such that $1/p+1/q=1$. Let $\eta_1,\eta_2\in L^1\cap L^q$ be densities of equal mass, i.e. $\int_{\mathbb{R}^d}\eta_1=\int_{\mathbb{R}^d}\eta_2$, and let $\sigma\in\Pi(\eta_1,\eta_2)$ be a coupling with marginals η_1 and η_2. Then for any integrable function u with $\nabla u\in L^p$, it holds*

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d}\frac{|u(x)-u(y)|}{|x-y|}\,\mathrm{d}\sigma(x,y)\lesssim(\|\eta_1\|_{L^q}+\|\eta_2\|_{L^q})\|\nabla u\|_{L^p}. \tag{3.24}$$

Proof. First, the statement is trivial if $p=\infty$ since u is Lipschitz and we arrive straightforwardly to the result of the lemma. Now, let $p\in(1,\infty)$. By means of the Morrey-type inequality (4.30) and using the marginal conditions for the measure $\sigma(x,y)$ we obtain

$$\begin{aligned}\iint_{\mathbb{R}^d\times\mathbb{R}^d}\frac{|u(x)-u(y)|}{|x-y|}\,\mathrm{d}\sigma(x,y)&\lesssim\iint_{\mathbb{R}^d\times\mathbb{R}^d}(M(\nabla u)(x)+M(\nabla u)(y))\,\mathrm{d}\sigma(x,y)\\&=\int_{\mathbb{R}^d}M(\nabla u)(x)\eta_1(x)\,\mathrm{d}x+\int_{\mathbb{R}^d}M(\nabla u)(y)\eta_2(y)\,\mathrm{d}y.\end{aligned}$$

Then by Hölder inequality and by the continuity of the maximal operator from L^p to L^p we deduce the statement of the Lemma. ■

Observe that this result can only be applied when $\nabla u \in L^p$ with $p > 1$. The limit case where ∇u is a singular integral of an L^1 function, that will be of interest in the next section, has to be dealt more carefully and requires some more elaborate tools from the Calderón–Zygmund theory.

Proof of Theorem 3.1. We consider the dual setting of the Kantorovich–Rubinstein distance between the solutions θ_1 and θ_2. By Lemma 2.2 we can write and set

$$\begin{aligned}\frac{d}{dt}\mathcal{D}_\delta(\theta_1,\theta_2) &= \int_{\mathbb{R}^d} \nabla\zeta_t \cdot (u_1\theta_1 - u_2\theta_2)\,\mathrm{d}x - \int_{\mathbb{R}^d} \nabla\zeta_t \cdot (\kappa_1\nabla\theta_1 - \kappa_2\nabla\theta_2)\,\mathrm{d}x \\ &= \int_{\mathbb{R}^d} \nabla\zeta_t \cdot (u_1\theta_1 - u_2\theta_2)\,\mathrm{d}x - \kappa_1\int_{\mathbb{R}^d} \nabla\zeta_t \cdot (\nabla\theta_1 - \nabla\theta_2)\,\mathrm{d}x \\ &\quad + (\kappa_2 - \kappa_1)\int_{\mathbb{R}^d} \nabla\zeta_t \cdot \nabla\theta_2\,\mathrm{d}x \\ &=: \Theta_1(t) + \Theta_2(t) + \Theta_3(t). \end{aligned} \tag{3.25}$$

We prove that the individual terms are controlled as follows

$$\Theta_1(t) \lesssim \|\nabla u_1\|_{L^p} + \frac{\|u_1 - u_2\|_{L^p}}{\delta}; \tag{3.26}$$

$$\Theta_2(t) \le 0; \tag{3.27}$$

$$\Theta_3(t) \le \frac{|\kappa_2 - \kappa_1|}{\delta}\|\nabla\theta_2\|_{L^1}. \tag{3.28}$$

Before turning to the proofs of these bounds, we can straightforwardly conclude the proof of Theorem 3.1. Indeed, inserting the bounds (3.26), (3.27) and (3.28) into (3.25) imply after integrating over $(0,t)$ for any $0 \le t \le T$ the stability estimate

$$\begin{aligned}\mathcal{D}_\delta(\theta_1(t,\cdot),\theta_2(t,\cdot)) - \mathcal{D}_\delta(\theta_1(0,\cdot),\theta_2(0,\cdot)) \lesssim \|\nabla u_1\|_{L^1(L^p)} + \frac{\|u_1 - u_2\|_{L^1(L^p)}}{\delta} \\ + \frac{|\kappa_1 - \kappa_2|}{\delta}\|\nabla\theta_2\|_{L^1},\end{aligned}$$

which is what we aimed to prove.

Proof of Estimate (3.26). For the first term $\Theta_1(t)$ we will use the dual representation of the optimal transportation distance. Using the properties of the Kantorovich potential

ζ_t and the marginal conditions of the optimal transport plan, it holds

$$
\begin{aligned}
\Theta_1(t) &= \int_{\mathbb{R}^d} \nabla\zeta_t \cdot (u_1\theta_1 - u_2\theta_2)\,\mathrm{d}x \\
&= \iint_{\mathbb{R}^d\times\mathbb{R}^d} (\nabla\zeta_t(x)\cdot u_1(x) - \nabla\zeta_t(y)\cdot u_2(y))\,\mathrm{d}\pi_t(x,y) \\
&= \iint_{\mathbb{R}^d\times\mathbb{R}^d} \frac{1}{|x-y|+\delta}\frac{x-y}{|x-y|}\cdot(u_1(x)-u_2(y))\,\mathrm{d}\pi_t(x,y),
\end{aligned}
$$

where π_t is the optimal transport plan in $\Pi((\theta_1(t)-\theta_2(t))^+,(\theta_1(t)-\theta_2(t))^-)$. We now separate the gradient term from the error term,

$$
|\Theta_1(t)| \le \iint_{\mathbb{R}^d\times\mathbb{R}^d} \frac{|u_1(x)-u_1(y)|}{|x-y|}\,\mathrm{d}\pi_t(x,y) + \frac{1}{\delta}\iint_{\mathbb{R}^d\times\mathbb{R}^d} |u_1(y)-u_2(y)|\,\mathrm{d}\pi_t(x,y).
$$

For the difference quotients in the first term, we apply Lemma 3.1. Regarding the second term, we can use the marginal conditions and the Hölder inequality,

$$
\frac{1}{\delta}\iint_{\mathbb{R}^d\times\mathbb{R}^d} |u_1(y)-u_2(y)|\,\mathrm{d}\pi_t(x,y) = \frac{1}{\delta}\int_{\mathbb{R}^d} |u_1-u_2|(\theta_1-\theta_2)^-\,\mathrm{d}y \le \frac{\|u_1-u_2\|_{L^p}}{\delta}\|\theta_2\|_{L^q}.
$$

All in all, we have established (3.26).

Proof of Estimate (3.27). The control of $\Theta_2(t)$, the second term in (3.25), is based on a discretization approach and hence we introduce *finite difference quotients*. Assume $v:\mathbb{R}^d\to\mathbb{R}$ is a locally summable function, then the i^{th} difference quotient with $1\le i\le d$ of size $h>0$ at $x\in\mathbb{R}^d$ is given by

$$
D_i^h v(x) = \frac{v(x+he_i)-v(x)}{h}. \tag{3.29}
$$

We also make use of these quotients to approximate the Laplacian by the standard three point stencil

$$
\Delta^h v(t,x) := \sum_{i=1}^d -D_i^{-h}D_i^v\theta(t,x) = \frac{1}{h^2}\sum_{i=1}^d (v(x+he_i)-2v(x)+v(x-he_i)).
$$

We follow a technique inspired from [46] consisting of a convenient rearrangement of the terms involved thanks to a discretization of the spatial derivatives. Let us take $h>0$ sufficiently small and write the finite differences approximation of the Laplacian using the difference quotients to arrive at a discretized analog of Θ_2 given for all $t \in (0,T)$ by

$$\begin{aligned}\frac{1}{\kappa_1}\Theta_2^h(t) &= \int_{\mathbb{R}^d} \zeta_t(x)\ \Delta^h\theta_1(t,x)\,\mathrm{d}x - \int_{\mathbb{R}^d} \zeta_t(x)\ \Delta^h\theta_2(t,x)\,\mathrm{d}x \\ &= \frac{1}{h^2}\int_{\mathbb{R}^d} \zeta_t(x)\sum_{i=1}^d \big(\theta_1(t,x+he_i) - 2\theta_1(t,x) + \theta_1(t,x-he_i)\big)\,\mathrm{d}x \\ &\quad - \frac{1}{h^2}\int_{\mathbb{R}^d} \zeta_t(x)\sum_{i=1}^d \big(\theta_2(t,x+he_i) - 2\theta_2(t,x) + \theta_2(t,x-he_i)\big)\,\mathrm{d}x.\end{aligned}$$

Since $\theta_1, \theta_2 \in W^{1,1}$ and $\zeta \in W^{1,\infty}$, we have that

$$\Theta_2(t) = \lim_{h\to 0} \Theta_2^h(t),$$

and thus, it is enough to estimate Θ_2^h instead of Θ_2.

In order to estimate Θ_2^h, we apply a convenient changes of variables $z_i = x + he_i$ and $y_i = x - he_i$ for all $1 \leq i \leq d$ to arrive at

$$\frac{1}{\kappa_1}\Theta_2^h(t) = \frac{1}{h^2}\int_{\mathbb{R}^d} (\theta_1(t,x) - \theta_2(t,x))\sum_{i=1}^d \big(\zeta_t(x+he_i) - 2\zeta_t(x) + \zeta_t(x-he_i)\big)\,\mathrm{d}x,$$

and exploring to the optimality of ζ_t therefore now ζ_{t,z_i} and ζ_{t,y_i} in the dual formulation of the Kantorovich–Rubinstein distance $\mathcal{D}_\delta(\theta_1,\theta_2)$, we obtain

$$\frac{h^2}{\kappa_1}\Theta_2^h(t) \leq -2d\mathcal{D}_\delta(\theta_1,\theta_2) + d\mathcal{D}_\delta(\theta_1,\theta_2) + d\mathcal{D}_\delta(\theta_1,\theta_2) = 0,$$

which proves (3.27).

Proof of Estimate (3.28). Finally, for $\Theta_3(t)$ we can use again that ζ_t is a Lipschitz function with Lipschitz constant bounded by $\|\nabla\zeta_t\|_{L^\infty} \leq 1/\delta$, so we arrive at (3.28). ■

3.4. Uniqueness with vector fields whose gradient is a singular integral of an integrable function

In this section, we are going to deal with the existence and uniqueness problem for the advection-diffusion equation (3.1) stated in Theorem 3.2. Before turning to its proof, we briefly verify that the velocity fields considered here belong indeed to a weak L^p space, globally in space. Along this section, we will denote by $\mathcal{L}^d$ the $d-$dimensional Lebesgue measure.

Lemma 3.2. *Let u be given by* (3.17) *with $\omega \in L^1(\mathbb{R}^d)$ and k satisfying* (k$_1$)–(k$_4$), *then it holds that*

$$u \in L^{p,\infty}(\mathbb{R}^d) \qquad \text{for } p = \frac{d}{d-1}.$$

Our assumption in (3.19) is a little bit stronger than what is implied by (3.18), but it falls into the class of velocity fields that are, for instance, induced by $L^\infty(L^1)$ vorticity solutions to the Euler equations as considered in [31].

The result is a classical result from harmonic analysis. We provide a short elementary proof for the convenience of the reader.

Proof. By rescaling ω, we may without loss of generality assume that $\|\omega\|_{L^1} = 1$. In view of the assumptions on the velocity field, we have the pointwise estimate

$$|u(x)| \lesssim \int_{\mathbb{R}^d} \frac{|\omega(y)|}{|x-y|^{d-1}} \,\mathrm{d}y.$$

Given some radius $R > 0$ that we will fix later, we now decompose the integral into a bounded and an integrable part,

$$\begin{aligned} |u(x)| &\lesssim \int_{B_R(x)} \frac{|\omega(y)|}{|x-y|^{d-1}} \,\mathrm{d}y + \int_{B_R(x)^c} \frac{|\omega(y)|}{|x-y|^{d-1}} \,\mathrm{d}y \\ &\leq \left(\chi_{B_R(0)} \frac{1}{|\cdot|^{d-1}} \right) * |\omega|(x) + \frac{1}{R^{d-1}}. \end{aligned} \tag{3.30}$$

An integral bound on the first term is obtained via an elementary computation that is based on Young's inequality,

$$\left\| \left(\chi_{B_R(0)} \frac{1}{|\cdot|^{d-1}} \right) * \omega \right\|_{L^1} \leq \left\| \chi_{B_R(0)} \frac{1}{|\cdot|^{d-1}} \right\|_{L^1} \|\omega\|_{L^1} \lesssim R. \tag{3.31}$$

Here, we do not keep track of the dependence of constants on ω.

In order to estimate the weak L^p norm, we let λ be arbitrarily given and suppose that $\lambda < |u(x)|$. Then we have by (3.30) for some R such that $\lambda \sim R^{1-d}$,

$$c\lambda \leq \left(\chi_{B_R(0)} \frac{1}{|\cdot|^{d-1}} \right) * |\omega|(x),$$

for some small c, and thus,

$$\mathcal{L}^{d+1}(\{|u| > \lambda\}) \leq \mathcal{L}^{d+1}\left(\left\{ \chi_{B_R(0)} \frac{1}{|\cdot|^{d-1}} * |\omega| > c\lambda \right\} \right) \lesssim \frac{1}{\lambda} \left\| \left(\chi_{B_R(0)} \frac{1}{|\cdot|^{d-1}} \right) * |\omega| \right\|_{L^1}.$$

As a consequence of (3.31) and the above choice of R, we deduce that

$$\lambda^p \mathcal{L}^{d+1}(\{|u| > \lambda\}) \lesssim \lambda^{p-1-\frac{1}{d-1}}.$$

Choosing p as in the statement of the lemma, we see that the right-hand side is in fact independent of λ and so is the supremum in λ, which yields the desired bound. ■

Towards a proof of the uniqueness result from Theorem 3.2, we will have to establish a suitably adapted version of the stability estimate in Theorem 3.1.

Proposition 3.1. *Under the assumptions of Theorem 3.2, let* $\theta \in L^\infty((0,T);(L^1 \cap L^\infty)(\mathbb{R}^d))$ *be a solution of* (3.1) *with* $\int_{\mathbb{R}^d} \theta^0 = 0$ *and* $\theta \not\equiv 0$*. Then for every* $\varepsilon > 0$ *there exists a constant* $C_\varepsilon > 0$ *such that for every* $\delta > 0$ *it holds*

$$\sup_{0\le t\le T} \mathcal{D}_\delta(\theta(t,\cdot)) \lesssim \mathcal{D}_\delta(\theta^0) + \varepsilon\|\theta\|_{L^1}\left[1 + \log\left(\frac{1}{\varepsilon\delta}\left(\frac{\|\theta\|_{L^1}}{\|\theta\|_{L^\infty}}\right)^{1-\frac{1}{p}}\|u\|_{L^{p,\infty}}\right)\right] + C_\varepsilon\|\theta\|_{L^\infty(L^2)}.$$

This estimate was derived earlier in the non-diffusive setting [31]. As the argument in the present work is essentially a combination of the one therein and the one that we proposed in order to establish Theorem 3.1, we will keep our presentation here short.

Regarding the proof, the main difference between the DiPerna–Lions setting considered in the previous section and the one we deal with here is the failure of the maximal function estimates. Instead of estimating difference quotients with the help of the Morrey-type estimate in (4.30), the strategy here is to construct certain weighted maximal functions which allow for the substitutive estimate

$$\frac{|u(t,x) - u(t,y)|}{|x-y|} \lesssim G(t,x) + G(t,y) \qquad \text{for all } x, y \notin N_t, \tag{3.32}$$

where N_t is a negligible set, $\mathcal{L}^d(N_t) = 0$, which exists for almost every time t, and $G : (0,T)\times\mathbb{R}^d \to \mathbb{R}$ is a function which can be decomposed for every $\varepsilon > 0$ into a sum $G = G^1_\varepsilon + G^2_\varepsilon$, where G^1_ε and G^2_ε are such that

$$\|G^1_\varepsilon\|_{L^1(L^{1,\infty})} \le \varepsilon, \qquad \|G^2_\varepsilon\|_{L^1(L^2)} \le C_\varepsilon. \tag{3.33}$$

If $\omega \in L^1(L^1)$, it is proved in [15] that such a function exists and the constant C_ε depends not only on $\varepsilon > 0$ but also on the equi-integrability of ω. Therefore, this result would not generalize to situations in which ω is simply a measure.

We will not give any details about the construction of the function G in (3.32) and (3.33). However, for the convenience of the reader, we provide here the full argument for Proposition 3.1.

Proof of Propositon 3.1. As we are seeking the stability estimate analogous to the one in Theorem 3.1, we start by computing the time derivative of the optimal transportation distance. Denoting ζ_t the Kantorovich potential at time t, by Lemma 2.2 we have

$$\frac{d}{dt}\mathcal{D}_\delta(\theta(t,\cdot)) = \int_{\mathbb{R}^d} \nabla\zeta_t \cdot u(t,x)\theta(t,x)\,\mathrm{d}x - \kappa\int_{\mathbb{R}^d} \nabla\zeta_t \cdot \nabla\theta(t,x)\,\mathrm{d}x.$$

Hence, by a similar analysis to the performed in Theorem 3.1, we obtain the following estimate

$$\sup_{t\in(0,T)} \mathcal{D}_\delta(\theta(t,\cdot)) \le \mathcal{D}_\delta(\theta^0) + \int_0^T \iint_{\mathbb{R}^d\times\mathbb{R}^d} \frac{|u(t,x)-u(t,y)|}{|x-y|+\delta}\,\mathrm{d}\pi_t(x,t). \tag{3.34}$$

In this case we cannot apply the Morrey inequality (4.30) as before. Nonetheless, we might make use of the alternative theory developed for weak L^p spaces and apply the inequalities (3.32) and (3.33). Therefore we can estimate pointwise the integrand in (3.34) for almost every $t\in(0,T)$ and every $x,y\in N_t$, where $\mathcal{L}^d(N_t)=0$ by

$$\begin{aligned}\frac{|u(t,x)-u(t,y)|}{|x-y|+\delta} \lesssim \min\left\{\frac{|u(t,x)|+|u(t,y)|}{\delta}, G^1_\varepsilon(t,x)+G^1_\varepsilon(t,y)\right\} \\ + G^2_\varepsilon(t,x)+G^2_\varepsilon(t,y).\end{aligned} \tag{3.35}$$

Notice that, since the marginals of the optimal transport plan π_t are absolutely continuous with respect to $\mathcal{L}^d$, the pointwise estimate holds for almost every $t\in(0,T)$ and π_t-almost every $(x,y)\in\mathbb{R}^d\times\mathbb{R}^d$.

To begin with, we take care of the first term in the right-hand side. We introduce this first part of the estimate into (3.34) and split the terms as follows,

$$\int_0^T \iint_{\mathbb{R}^d\times\mathbb{R}^d} \min\left\{\frac{|u(t,x)|+|u(t,y)|}{\delta}, G^1_\varepsilon(t,x)+G^1_\varepsilon(t,y)\right\}\mathrm{d}\pi_t(x,y)\,\mathrm{d}t \le I_1+I_2,$$

where

$$\begin{aligned}I_1 &= \int_0^T \iint_{\mathbb{R}^d\times\mathbb{R}^d} \left(\min\left\{\frac{|u(t,x)|}{\delta}, G^1_\varepsilon(t,x)\right\} + \min\left\{\frac{|u(t,y)|}{\delta}, G^1_\varepsilon(t,y)\right\}\right)\mathrm{d}\pi_t(x,y)\,\mathrm{d}t,\\ I_2 &= \int_0^T \iint_{\mathbb{R}^d\times\mathbb{R}^d} \left(\min\left\{\frac{|u(t,x)|}{\delta}, G^1_\varepsilon(t,y)\right\} + \min\left\{\frac{|u(t,y)|}{\delta}, G^1_\varepsilon(t,x)\right\}\right)\mathrm{d}\pi_t(x,y)\,\mathrm{d}t.\end{aligned}$$

By means of the marginal condition for the optimal transport plan we have

$$I_1 = \int_0^T \int_{\mathbb{R}^d} \min\left\{\frac{|u(t,x)|}{\delta}, G_\varepsilon^1(t,x)\right\} |\theta(t,x)| \,\mathrm{d}x\,\mathrm{d}t.$$

In order to make the notation simpler, we write $\psi = \min\{|u|/\delta, G_\varepsilon^1\}$. The main challenge now comes from the fact that G_ε^1 can only be bounded in the weak space $L^{1,\infty}$ while u in $L^{p,\infty}$. Let us define the finite measure

$$\mathrm{d}\mu(t,x) = \chi_{(0,T)}(t)|\theta(t,x)| \;\mathrm{d}\big(\mathcal{L}^1 \otimes \mathcal{L}^d\big)$$

in $\mathbb{R}^{d+1}$ so that, since ψ is defined as a minimum, we can bound on the one hand

$$\|\psi\|_{L^{1,\infty}(\mu)} \le \|G_\varepsilon^1\|_{L^{1,\infty}(\mu)} \le \|\theta\|_{L^\infty} \|G_\varepsilon^1\|_{L^1(L^{1,\infty})} \le \varepsilon\|\theta\|_{L^\infty} \tag{3.36}$$

and on the other hand by using (3.33) also

$$\|\psi\|_{L^{p,\infty}(\mu)} \le \frac{1}{\delta}\|u\|_{L^{p,\infty}(\mu)} \le \frac{1}{\delta}\|\theta\|_{L^\infty}^{1/p}\|u\|_{L^{p,\infty}}. \tag{3.37}$$

Now, using from Lemma 2.6 in [31] the interpolation inequality

$$\|\psi\|_{L^1(\mu)} \le \frac{p}{p-1}\|\psi\|_{L^{1,\infty}(\mu)} \left[1 + \log\left(\frac{\mu(\mathbb{R}^{d+1})^{1-\frac{1}{p}}\|\psi\|_{L^{p,\infty}(\mu)}}{\|\psi\|_{L^{1,\infty}(\mu)}}\right)\right],$$

and the monotonicity of the expression on the right-hand side in the $L^{1,\infty}$-norm, we find for any $\theta \not\equiv 0$ the bound

$$I_1 = \|\psi\|_{L^1(\mu)} \lesssim \varepsilon\|\theta\|_{L^\infty} \left[1 + \log\left(\frac{1}{\varepsilon\delta}\left(\frac{\|\theta\|_{L^1}}{\|\theta\|_{L^\infty}}\right)^{1-\frac{1}{p}}\|u\|_{L^{p,\infty}}\right)\right]. \tag{3.38}$$

The argument for I_2 will be similar to the just performed estimates. To do so, we need to prove that the estimates (3.36) and (3.37) hold also in the situation of I_2. Recall the characterization of the optimal transport plans through the measurable maps T and S, introduced in Section 2.2, which together with the marginal condition give

$$\begin{aligned} I_2 = &\int_0^T \int_{\mathbb{R}^d} \min\left\{\frac{|u\circ S|(t,y)}{\delta}, G_\varepsilon^1(t,y)\right\} \theta^-(t,y)\,\mathrm{d}y\,\mathrm{d}t \\ &+ \int_0^T \int_{\mathbb{R}^d} \min\left\{\frac{|u\circ T|(t,x)}{\delta}, G_\varepsilon^1(t,x)\right\} \theta^+(t,x)\,\mathrm{d}x\,\mathrm{d}t. \end{aligned}$$

The treatment of the two terms now is quite similar, therefore it would be enough to focus on one of them, say the last one. Analogously to the estimate for I_1, we define now the function

$$\psi = \min\left\{\frac{|u\circ T|}{\delta}, G^1_\varepsilon\right\},$$

and the finite measure $d\mu(t,x) = \chi_{(0,T)}(t)\theta^+(t,x)d\mathcal{L}^1\otimes\mathcal{L}^d$ on $\mathbb{R}^{d+1}$. The first estimate (3.36) remains valid without any change since it comes from assuming $\psi \le G^1_\varepsilon$, that holds true. For the second estimate (3.37), however, we need to take care of some details. On the one hand, we have

$$\|\psi\|_{L^{p,\infty}(\mu)} \le \frac{1}{\delta}\|u\circ T\|_{L^{p,\infty}(\mu)},$$

and on the other hand, we have the relation $\theta^- = T_\#\theta^+$, which implies

$$\mu(\{|u\circ T|>\lambda\}) = \left(T_\#\theta^+\mathcal{L}^1\otimes\mathcal{L}^d\right)(\{|u|>\lambda\}) = \left(\theta^-\mathcal{L}^1\otimes\mathcal{L}^d\right)(\{|u|>\lambda\}).$$

Therefore, we have

$$\begin{aligned}\|u\circ T\|_{L^{p,\infty}(\mu)} &= \sup_{\lambda>0}\left(\lambda^p\int_0^T\int_{\mathbb{R}^d}\chi_{\{|u\circ T|>\lambda\}}\theta^+(t,x)\,\mathrm{d}x\,\mathrm{d}t\right)^{1/p}\\ &= \sup_{\lambda>0}\left(\lambda^p\int_0^T\int_{\mathbb{R}^d}\chi_{\{|u|>\lambda\}}\theta^-(t,x)\,\mathrm{d}x\,\mathrm{d}t\right)^{1/p} \le \|\theta\|^{1/p}_{L^\infty}\|u\|_{L^{p,\infty}}.\end{aligned}$$

Hence the estimate (3.37) also holds for I_2 and we arrive to the estimate (3.38) in this case as well.

Finally, we can control the terms related to G^2_ε in (3.35) by means of the marginal conditions for the optimal transport plan and Hölder inequality,

$$\int_0^T\int_{\mathbb{R}^d}(G^2_\varepsilon(t,x)+G^2_\varepsilon(t,y))\,\mathrm{d}\pi_t(x,y) = \int_0^T\int_{\mathbb{R}^d}G^2_\varepsilon|\theta|\,\mathrm{d}t\,\mathrm{d}x \le \|\theta\|_{L^\infty(L^2)}\|G^2_\varepsilon\|_{L^1(L^2)}. \tag{3.39}$$

that is bounded since $\theta\in L^\infty(L^2)$ by interpolation. Therefore we can plug the estimates (3.38) and (3.39) into (3.35) and it yields the desired stability estimate. ■

From Proposition 3.1 it is easy to deduce uniqueness with the help of Lemma 2.1. It remains to prepare for the proof of existence. We will establish existence by a standard mollification-and-compactness procedure. For this, we provide an auxiliary lemma about the convergence of the velocity fields in appropriate Lebesgue spaces that we present here in a quantitative way. We believe this result is of independent mathematical interest.

Lemma 3.3. *Consider $\omega_n, \omega : \mathbb{R}^d \to \mathbb{R}^d$ integrable functions all with the same total mass. Assume that $\|\omega_n - \omega\|_{L^1(\mathbb{R}^d)}$ is bounded uniformly in $n \in \mathbb{N}$. Let $u_n = k * \omega_n$, $u = k * \omega$ with k satisfying* (k$_1$)–(k$_4$)*, then for every $s > 0$ and any $1 \le p < d/(d-1)$ it holds*

$$\|u_n - u\|_{L^p(B_s(0))} \lesssim \left(s^{\frac{d}{p}} W_1(\omega_n, \omega)\right)^{\frac{d-p(d-1)}{d+p}}.$$

This lemma states that under a suitable convergence assumptions for ω_n in $L^1(\mathbb{R}^d)$ we can control the $L^p_{loc}(\mathbb{R}^d)$ convergence of u_n in terms of the Wasserstein distance for $1 \le p < d/(d-1)$. The most evident consequence of this lemma is the following result, whose proof is now obvious.

Lemma 3.4. *Under the same assumptions of Lemma 3.3, if $W_1(\omega_n, \omega) \to 0$ as $n \to \infty$, then u_n converges to u locally in $L^p(\mathbb{R}^d)$, provided that $1 \le p < d/(d-1)$.*

Proof of Lemma 3.3. First, we show that $\omega \in L^1(\mathbb{R}^d)$ implies $u \in L^p_{loc}(\mathbb{R}^d)$. For this purpose, it is convenient and enough to assume that ω is nonnegative density of mass one, $\|\omega\|_{L^1} = 1$. We choose $s > 0$ arbitrary, and derive by means of the Jensen inequality applied with the measure $\omega(y)dy$ and Fubini's theorem,

$$\begin{aligned}\int_{B_s(0)} |u|^p \,\mathrm{d}x &\lesssim \int_{B_s(0)} \left(\int_{\mathbb{R}^d} \frac{\omega(y)}{|x-y|^{d-1}} dy\right)^p \mathrm{d}x \lesssim \int_{B_s(0)} \int_{\mathbb{R}^d} \frac{\omega(y)}{|x-y|^{p(d-1)}} \,\mathrm{d}y \,\mathrm{d}x \\ &= \int_{\mathbb{R}^d} \left(\int_{B_s(0)} \frac{1}{|x-y|^{p(d-1)}} \,\mathrm{d}x\right) \omega(y) \,\mathrm{d}y \lesssim s^{d-p(d-1)}\end{aligned}$$

provided that $1 \le p < d/(d-1)$. In order to prove the strong convergence of $(u_n)_{n\in\mathbb{N}}$ to u in $L^p_{loc}(\mathbb{R}^d)$ we first study the pointwise difference $u_n(x) - u(x)$. Let $R > 0$ be such that $k(z)$ is Lipschitz for all $z \in B_R(0)^c$, and define the cutoff functions $\eta_R : \mathbb{R}^d \to [0,1]$, $\eta_R \in C_c^\infty(\mathbb{R}^d)$, $\eta_R(x) = 1$ for all $x \in B_R(0)$ and $\eta_R(x) = 0$ for all $x \in \overline{B_{2R}(0)}^c$. Then $\varphi(x) = (1-\eta_R)k(x)$ is a Lipschitz function with Lipschitz constant bounded by $\|\nabla\varphi\|_{L^\infty(\mathbb{R}^d)} \lesssim R^{-d}$. Thus, we can write

$$\begin{aligned}u_n(x) - u(x) &= \int_{\mathbb{R}^d} k(x-y)[\omega_n(y) - \omega(y)] \,\mathrm{d}y \\ &= \int_{\mathbb{R}^d} (\eta_R k)(x-y)[\omega_n(y) - \omega(y)] \,\mathrm{d}y + \int_{\mathbb{R}^d} \varphi(x-y)[\omega_n(y) - \omega(y)] \,\mathrm{d}y.\end{aligned}$$

Since $\varphi(x)$ is Lipschitz, the second term in the right hand side can be related to the 1-Wasserstein distance (2.8), provided ω_n and ω are of the same total mass,

$$\left|\int_{\mathbb{R}^d} \varphi(x-y)[\omega_n(y) - \omega(y)] \,\mathrm{d}y\right| \lesssim \frac{1}{R^d} W_1(\omega_n, \omega).$$

On the other hand, the remaining term can be bounded by

$$\int_{\mathbb{R}^d} (\eta_R k)(x-y)[\omega_n(y)-\omega(y)]\,\mathrm{d}y \lesssim \int_{B_{2R}(x)} \frac{1}{|x-y|^{d-1}} |\omega_n(y)-\omega(y)|\,\mathrm{d}y = (F_{2R} * |\omega_n - \omega|)(x),$$

where the function F_r is defined as $F_r(z) = |z|^{1-d}\chi_{B_r(0)}(z)$ for any $r > 0$. Thus, for every $s > 0$ it holds

$$\|u_n - u\|_{L^p(B_s(0))} \lesssim \|F_{2R} * |\omega_n - \omega|\|_{L^p(\mathbb{R}^d)} + \|R^{-d} W_1(\omega_n, \omega)\|_{L^p(B_s(0))}.$$

For the first term, we might use the Young's convolution inequality so that we get

$$\|F_{2R} * |\omega_n - \omega|\|_{L^p(\mathbb{R}^d)} \leq \|F_{2R}\|_{L^p(\mathbb{R}^d)} \|\omega_n - \omega\|_{L^1(\mathbb{R}^d)} \lesssim R^{\frac{d-p(d-1)}{p}}$$

given that $\|\omega_n - \omega\|_{L^1(\mathbb{R}^d)}$ is bounded uniformly in $n \in \mathbb{N}$ and that $1 \leq p < d/(d-1)$. The second term on the right-hand side is the norm of a constant, thus

$$\|u_n - u\|_{L^p(B_s(0))} \lesssim R^{\frac{d-p(d-1)}{p}} + s^{\frac{d}{p}} R^{-d} W_1(\omega_n, \omega).$$

But we can optimize the bound in $R > 0$ to the effect that for all $s > 0$ it holds

$$\|u_n - u\|_{L^p(B_s(0))} \lesssim \left(s^{\frac{d}{p}} W_1(\omega_n, \omega)\right)^{\frac{d-p(d-1)}{d+p}}. \qquad \blacksquare$$

We are now in the position to prove Theorem 3.2. In order to do so we will rely on the hyperbolic tangent distance $\mathcal{D}^b$ and Lemma 2.1 from Section 2.2, that is defined from the cost function

$$c(z) = \tanh(z).$$

Proof of Theorem 3.2. First of all, we want to give a sketch of an existence proof. On this regard, we notice that distributional solutions are well-defined because $\nabla\theta \in L^1((0,T) \times \mathbb{R}^d)$, see Remark 3.1, and because $u\theta \in L^1((0,T) \times \mathbb{R}^d)$. The letter follows from the estimate

$$\|u\theta\|_{L^1} = \|u\|_{L^1(\mu)} \leq \frac{p}{p-1} \|\theta\|_{L^1}^{1-\frac{1}{p}} \|u\|_{L^{p,\infty}(\mu)} \leq \frac{p}{p-1} \|\theta\|_{L^1}^{1-\frac{1}{p}} \|\theta\|_{L^\infty}^{\frac{1}{p}} \|u\|_{L^{p,\infty}} < +\infty, \tag{3.40}$$

where the measure μ is defined by $\mu(t,x) = \chi_{(0,T)}(t)\theta(x)\mathcal{L}^1 \otimes \mathcal{L}^d$ — we assume that θ is nonnegative for convenience — and the first inequality is due to the embedding $L^{p,\infty}(X,\mu) \subset L^1(X,\mu)$ on a finite measure space (X,μ), see, e.g., Lemma 2.5 in [31].

Now, to prove existence, we will proceed by regularizing the velocity field and initial datum and then passing to the limit under the appropriate conditions. Denoting by ρ_ε a standard mollifier on $\mathbb{R}^d$, we define $\omega_\varepsilon = \omega * \rho_\varepsilon \in L^1((0,T); C_b^1(\mathbb{R}^d))$ and $\theta_\varepsilon^0 = \theta^0 * \rho_\varepsilon \in C_b^1(\mathbb{R}^d)$. Then, since $u = k * \omega$, we can also define $u_\varepsilon = k * \omega_\varepsilon \in L^1((0,T); C_b^1(\mathbb{R}^d))$. Therefore, by standard theory, we know that there exist a unique solution $\theta_\varepsilon \in C((0,T); C_b^1(\mathbb{R}^d))$ of the Cauchy problem

$$\begin{cases} \partial_t \theta_\varepsilon + \nabla \cdot (u_\varepsilon \theta_\varepsilon) = \kappa \Delta \theta_\varepsilon & \text{in } (0,T) \times \mathbb{R}^d, \\ \theta_\varepsilon(0,\cdot) = \theta_{0,\varepsilon} & \text{in } \mathbb{R}^d. \end{cases}$$

Now we can use the elementary a priori estimates (3.4) and our assumptions on the initial datum (3.12) in order to deduce that θ_ε is bounded in $L^\infty((0,T); L^q(\mathbb{R}^d))$ independently of $\varepsilon > 0$ for every $1 \le q \le \infty$. It follows that we can extract a subsequence (not relabelled) that converges weakly-$*$ to some function θ in $L^\infty(L^q)$ for any $q \in (1,\infty]$, and then by invoking some soft arguments, also in $L^\infty(L^1)$. Moreover, inspection of (3.1) reveals that the time derivatives $\partial_t \theta_\varepsilon$ are bounded in $L^\infty(H^{-s})$ for some $s > 0$, from which we infer that that the convergence takes place in $C^0(\text{w-}L^q)$ for any $q \in [1,\infty]$, where w-$L^q(\mathbb{R}^d)$ is the standard L^q space equipped with the weak topology. In view of Lemma 3.4 and (3.13) and because Wasserstein distances metrize weak convergence, see Theorem 7.12 in [94], the velocity fields u_ε are converging locally in any L^q space. As a consequence, the product $u_\varepsilon \theta_\varepsilon$ is convergent on compact sets, and thus, passing to the limit in the distributional formulation of (3.1), see Definition 2.1, we find that θ solves the advection-diffusion equation with velocity u and initial datum θ^0.

The proof of the uniqueness relies on the stability estimate from Proposition 3.1. Towards a contradiction, we assume that there is a solution $\theta(t,x)$ of the advection-diffusion equation (3.1) with initial datum $\theta^0 \equiv 0$ and such that $\theta(t,x) \not\equiv 0$, so that, in particular, $\|\theta\|_{L^\infty} > 0$. Then we can write

$$\sup_{0 \le t \le T} \mathcal{D}_\delta(\theta(t,\cdot)) \lesssim \varepsilon \left[1 + \log\left(\frac{1}{\varepsilon\delta}\right)\right] + C_\varepsilon,$$

where the symbol $\lesssim$ now includes $\|\theta\|_{L^\infty}$, $\|\theta\|_{L^1}$ and $\|u\|_{L^{p,\infty}}$. Notice that if $\delta \in (0, 1/e)$ it holds

$$\frac{1}{|\log \delta|}\left[1 + \log\left(\frac{1}{\delta\varepsilon}\right)\right] \le \frac{1 + |\log\delta| + |\log\varepsilon|}{|\log\delta|} \le 2 + |\log\varepsilon|,$$

and, therefore, we can choose $a > 0$ arbitrarily small and fix $\varepsilon > 0$ such that

$$\frac{\varepsilon}{|\log\delta|}\left[1 + \log\left(\frac{1}{\delta\varepsilon}\right)\right] \le \frac{a}{2}.$$

Since $\varepsilon > 0$ and $C_\varepsilon > 0$ are fixed now, we may choose $\delta \in (0, 1/e)$ small enough so that

$$\frac{C_\varepsilon}{|\log \delta|} \leq \frac{a}{2}.$$

Combining the previous estimates, we find that

$$\sup_{0 \leq t \leq T} \mathcal{D}_\delta(\theta(t, \cdot)) \lesssim a |\log \delta|.$$

Thus, since $a > 0$ was arbitrarily small, it holds

$$\frac{\mathcal{D}_\delta(\theta(t, \cdot))}{|\log \delta|} \to 0 \quad \text{as } \delta \to 0.$$

To conclude, it only remains to notice that Lemma 2.1 with $\gamma = \sqrt{\delta}$ implies that

$$\mathcal{D}^b(\theta(t, \cdot)) \leq 2 \frac{\mathcal{D}_\delta(\theta(t, \cdot))}{|\log \delta|} + \sqrt{\delta} \|\theta(t, \cdot)\|_{L^1(\mathbb{R}^d)}$$

for all $\delta > 0$ small enough. In particular letting $\delta \to 0$ we get $\mathcal{D}^b(\theta(t, \cdot)) = 0$ and since $\mathcal{D}^b(\cdot)$ is a norm, it implies $\theta \equiv 0$. This contradicts the hypothesis at the beginning of the proof and since we found that the only solution of (3.1) with initial datum $\theta^0 = 0$ is $\theta(t, x) = 0$ almost every $(t, x) \in (0, T) \times \mathbb{R}^d$, it yields the sought uniqueness. ■

4. Error estimates for a finite volume scheme with rough coefficients

This chapter is based on the article [71], which is a joint work with André Schlichting. Large parts of it are reproduced verbatim.

Chapter summary

We study the implicit upwind finite volume scheme for numerically approximating the advection-diffusion equation with a vector field in the low regularity DiPerna–Lions setting. That is, we are concerned with advecting velocity fields that are spatially Sobolev regular and data that are merely integrable. We prove that on unstructured regular meshes the rate of convergence of approximate solutions generated by the upwind scheme towards the unique solution of the continuous model is at least one. The numerical error is estimated in terms of logarithmic Kantorovich-Rubinstein distances and provides thus a bound on the rate of weak convergence.

4.1. Introduction

In this chapter we are concerned with a bounded domain $\mathrm{D} \subset \mathbb{R}^d$, a bounded time interval $(0,T)$ and a positive constant diffusion coefficient $\kappa > 0$. Given vector field $u : [0,T] \times \mathrm{D} \to \mathbb{R}^d$, we study the evolution of a scalar quantity $\theta : [0,T) \times \mathrm{D} \to \mathbb{R}$ described by the Cauchy problem

$$\begin{cases} \partial_t \theta + \nabla \cdot (u\theta) &= \kappa \Delta \theta \quad \text{in } (0,T) \times \mathrm{D}, \\ \theta(0,\cdot) &= \theta^0 \quad \text{in } \mathrm{D}, \end{cases} \tag{4.1}$$

where θ^0 is the initial configuration.

In addition we assume that there is no loss of mass across the boundary of the domain,

$$(\kappa \nabla \theta - u) \cdot n = 0 \quad \text{in } (0,T) \times \partial \mathrm{D}, \tag{4.2}$$

where $n = n(x)$ represents the outer unit vector normal to the boundary of the domain on every point $x \in \partial\mathrm{D}$. This assumption implies that solutions to the advection-diffusion equation (4.1) conserve their mass in time,

$$\int_\Omega \theta(t,x)\,\mathrm{d}x = \int_\Omega \theta^0(x)\,\mathrm{d}x \quad \text{for all } t \in (0,T).$$

In this context, well-posedness of renormalized solutions to the equation (4.1) is obtained for Sobolev regular vector fields by DiPerna and Lions [37]. This new solution concept is based on the least possible regularity such that the chain rule still holds, providing qualitative stability and hence uniqueness results. We say a vector field u is in the DiPerna–Lions setting if for some $1 < p \leq \infty$ it holds

$$u \in L^1((0,T); W^{1,p}(\mathrm{D})) \qquad \text{and} \qquad (\nabla \cdot u)^- \in L^1((0,T); L^\infty(\mathrm{D})). \tag{4.3}$$

For works explicitly handling diffusion in this regularity setting, see also [13, 45, **?**].

Considering then $\theta^0 \in L^q$ with $q > 1$ such that $1/p + 1/q \leq 1$, there is a unique distributional solution to the advection-diffusion equation (4.1) with vector field in the DiPerna–Lions setting such that

$$\theta \in L^\infty((0,T); L^q(\mathrm{D})) \cap L^1((0,T); W^{1,1}(\mathrm{D})).$$

Such regularity for the solution to (4.1) can be straightforwardly derived from the standard apriori estimate

$$\frac{1}{q(q-1)}\frac{d}{dt}\|\theta\|_{L^q}^q + \kappa \int_\mathrm{D} |\theta|^{q-2}|\nabla\theta|^2\,\mathrm{d}x \leq \frac{1}{q}\|(\nabla \cdot u)^-\|_{L^\infty}\|\theta\|_{L^q}^q, \tag{4.4}$$

with $q > 1$. Then one can see that the solution begin $L^\infty((0,T); L^q(\mathrm{D}))$ is obtained by integrating (4.4) and dropping the term with $\nabla\theta$ so that we get

$$\|\theta\|_{L^\infty(L^q)} \leq \Lambda^{1-\frac{1}{q}}\|\theta^0\|_{L^q}, \tag{4.5}$$

where $\Lambda = \exp(\|(\nabla \cdot u)^-\|_{L^1(L^\infty)})$ is the compressibility constant of the vector field. Since we dropped the term with $\nabla\theta$ in order to get (4.5), the estimate holds for both the transport equation ($\kappa = 0$) and the advection-diffusion equation ($\kappa > 0$). However, the presence of diffusion provides better regularity for the solution, which is obtained from the term involving $\nabla\theta$ in (4.4) as

$$\kappa \int_0^T \int_{\mathrm{D}} |\theta|^{q-2} |\nabla\theta|^2 \,\mathrm{d}x \,\mathrm{d}s \leq \frac{1}{q}\left(\frac{1}{q-1} + \Lambda^{q-1}\log\Lambda\right) \|\theta^0\|_{L^q}^q. \tag{4.6}$$

This control over the gradient provides that the solution to (4.1) lives in $L^1((0,T);W^{1,1}(\mathrm{D}))$ (see the beginning of Section 4.3).

The main objective of this chapter is to develop (optimal) error estimates for an upwind scheme on unstructured meshes based on a finite volume approximation of distributional solutions to the advection-diffusion equation (4.1) when the vector field is in the DiPerna–Lions setting. This result arises as a continuation of the works by Schlichting and Seis [79, 80], where the authors study the upwind scheme for the transport equation, i.e., $\kappa = 0$, in a similar regularity setting. The addition of a diffusive term is not trivial whatsoever, as we will explain in detail along the sections of this chapter. The main result in [72, Theorem 1] provides the stability estimate in the presence of diffusion, which will be a key ingredient for derivation of error estimates for the numerical scheme. In the DiPerna–Lions setting, the stability for two solutions is measured with respect to the optimal transport distance defined for any $\delta > 0$ by

$$\mathcal{D}_\delta(\mu_1,\mu_2) = \inf_{\pi\in\Pi(\mu_1,\mu_2)} \iint_{\mathrm{D}\times\mathrm{D}} \log\left(\frac{|x-y|}{\delta} + 1\right) \mathrm{d}\pi(x,y). \tag{4.7}$$

Here $\Pi(\mu_1,\mu_2)$ represents the set of all transport plans between the measures μ_1 and μ_2. We give a more in-depth contextualization and further explanation about these so-called *Kantorovich-Rubinstein distances* or optimal transport distances in Section 2.2.

The result [72, Theorem 1] states that any two solutions θ_1 and θ_2 of the advection-diffusion equation (4.1) with initial data θ_1^0, θ_2^0, vector fields u_1, u_2 and diffusion coefficients κ_1, κ_2 respectively, satisfy

$$\sup_{0\leq t\leq T} \mathcal{D}_\delta(\theta_1(t),\theta_2(t)) \lesssim \mathcal{D}_\delta(\theta_1^0,\theta_2^0) + 1 + \frac{\|u_1-u_2\|_{L^1(L^p)} + |\kappa_1-\kappa_2|\|\nabla\theta_2\|_{L^1}}{\delta}. \tag{4.8}$$

The study of convergence rates for finite volume schemes for the advection-diffusion equation is intimately related to the study of the diffusionless case, that was firstly addressed by Kuznetsov [56]. For results about the mathematical theory and the derivation of optimal error estimates with Lipschitz vector fields and regular initial data, either BV or H^1, see [34, 65]. On the DiPerna–Lions setting, the problem has not been addressed until very recently with the work of Schlichting and Seis, first with Cartesian meshes [79] and after with unstructured meshes [80].

The study of numerical approximations for the advection equation with diffusion has been of great interest along the last decades, from classical results with $C^1(\overline{\mathrm{D}})$ vector fields and Cartesian meshes [78] to less regular settings [93].

The novelty in our work is that we present error estimates for the finite volume scheme for the advection-diffusion in the low regularity framework. Denoting by h the size of the mesh and k to the time step, we get an $\mathcal{O}(h + \sqrt{k})$ error bound as in the smooth setting. We derive most of the results and estimates here working in the Eulerian setting for the equation (4.1), that is, operating with the solution of the partial differential equation. In previous works, mainly for the transport equation, the Lagrangian setting has been considered instead, i.e., the characteristics associated with the equation. This provides a probabilistic interpretation of the numerical scheme as a Markov chain on the mesh (see [35, 79]). In this chapter, we need to use the Lagrangian setting to prove one estimate related to the time-discretization of the vector field. Since we are dealing with a parabolic equation, the characteristics are solutions to stochastic differential equation, for which reason we include a short introduction to Lagrangian stochastic flows in Appendix A.1.

In addition, it is remarkable that working in a low regularity setting carries over a substantial change in topology compared to the smooth setting. Here we quantify the rate of *weak convergence*, following the spirit of previous works for the transport equation, e.g. [35, 79, 80]. For Lipschitz vector fields instead, it is possible to derive bounds in strong norms. However, for the DiPerna–Lions setting, we introduce the Kantorovich–Rubinstein distance that metrize weak convergence and hence it is a natural tool for studying this case, since only for those stability estimates are available [72].

In this work, we focus on the upwind finite volume scheme for linear advection, since it is the easiest to analyze and has the needed stability properties. An interesting question is, if the here presented proofs generalize to the analysis of structure preserving schemes for singular aggregation-diffusion equations, like the ones studied for regular aggregation in [36, 81].

This chapter is organized as follows: In Section 4.2 we present a precise definition of the admissible meshes, the finite volume numerical scheme, and its properties together with a presentation and a discussion of the main results. Section 4.3 contains all the proofs related to the main result of this chapter. Finally, Appendix A.1 provides an overview of stochastic Lagrangian flows on bounded domains, which is a needed tool to estimate the error related to the time-discretization of the vector field.

4.2. Setting and main result

4.2.1. Definition of the numerical scheme

In this section we present a formal and detailed definition of the upwind scheme that we will use. To begin with, recall from [43] the definition of admissible meshes for the finite volume discretization of advection-diffusion equations.

Definition 4.1 (Admissible meshes)**.** Let $\mathrm{D} \subset \mathbb{R}^d$ be an open, locally convex and bounded set with $C^{1,1}$ boundary. We say $\mathcal{T}$ is an admissible tessellation of D if it consists of a finite family of cells or control volumes $K \in \mathcal{T}$ and a finite family of points $\{x_K\}_{K\in\mathcal{T}} \subset \overline{\mathrm{D}}$ such that

- every control volume $K \in \mathcal{T}$ is a closed, connected and convex subset in D;
- the control volumes have disjoint interiors and satisfy $\overline{\mathrm{D}} = \bigcup_{K\in\mathcal{T}} K$;
- each cell is polygonal in the interior of D, in the sense that the interior boundary of each cell $\partial K \setminus \partial \mathrm{D}$ is the union of finitely many subsets of D contained in hyperplanes of $\mathbb{R}^d$ with strictly positive $\mathcal{H}^{d-1}$-measure;
- the family of points $\{x_K\}_{K\in\mathcal{T}}$ satisfies $x_K \in \overline{K} \setminus \partial\mathrm{D}$ for all $K \in \mathcal{T}$;.

In general, see [43], the geometry of $\partial\mathrm{D}$ is restricted to the case in which it is polygonal itself. However in our specific case, we need a construction of a *stochastic Lagrangian flow* (see Appendix A.1), for which certain error terms can only be controlled on domains satisfying a *uniform exterior ball condition* (4.11), for which a $C^{1,1}$ boundary is a sufficient condition. Since, we are working under a no-flow boundary condition (4.2), we can indeed consider sufficiently smooth domains D such that Definition 4.1 holds and the numerical cells are only polygonal inside of the domain D.

It is important to remark that the convexity requirement for the cells is needed in our analysis in order to prove Lemma 4.10 invoking a specific construction, the Brenier maps. Nonetheless we believe that this might not be strictly needed in general and one could come up with an similar construction that allows some relaxation for the convexity assumption.

A two dimensional example of two admissible control volumes is illustrated in Figure 4.1. We denote by $L \sim K$ whenever K and L are two neighbouring cells and we write $K|L$ to denote the common edge. If $L \sim K$, we define $d_{KL} = |x_L - x_K|$ and n_{KL} to be the unit vector on $K|L$ pointing in the direction $x_L - x_K$. In addition, abusing the notation, we write $|K|L| = \mathcal{H}^{d-1}(K|L)$ the $(d-1)$-dimensional Hausdorff measure of the edge $K \cap L$ and $|K| = \mathcal{L}^d(K)$ the d-dimensional Lebesgue measure of a cell $K \in \mathcal{T}$.

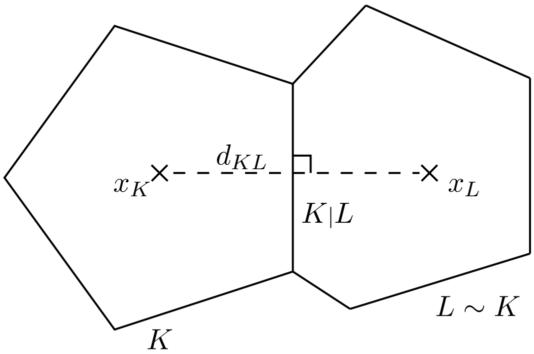

Figure 4.1.: Example of admissible neighbouring control volumes

The *mesh size* h is defined to be the maximal cell diameter, $h = \max_{K\in\mathcal{T}} \operatorname{diam} K$, and hence it holds $d_{KL} \lesssim h$ for all $K \in \mathcal{T}$ and $L \sim K$. For the time discretization we call k the *time step* such that there exists $N \in \mathbb{N}$ with $T = kN$ and we adopt the convention $t_j = jk$ for all $0 \le j \le N$. For the sake of shorter notation, we write $[\![0, N]\!]$ to denote the collection of numbers $\{0, 1, \ldots, N\}$.

In addition it is required to consider some regularity assumptions for the boundary of the domain and the mesh to ensure that, at least, the standard geometric constants arising on the Poincaré and the trace inequalities do not depend on the size of the mesh. Namely, it is needed that for every $f \in W^{1,1}(K) \cap C(\overline{K})$,

$$\begin{aligned} \|f\|_{L^1(\partial K)} &\lesssim \|\nabla f\|_{L^1(K)} + h^{-1}\|f\|_{L^1(K)}, \\ \|f - f_K\|_{L^1(K)} &\lesssim h\|\nabla f\|_{L^1(K)}, \end{aligned} \tag{4.9}$$

uniformly in $K \in \mathcal{T}$ and $h > 0$. These are respectively the trace and Poincaré inequalities and for a classical proof of these results we refer to [42, Sections 4.3 and 4.5]. We denote by f_K the average of f over the cell K, to be more specific $f_K = \fint_K f dx$. One direct consequence of the trace estimate is the so-called *isoperimetric property* of the mesh, that guarantees that every cell K of the tessellation has a volume of order h^d and a surface of order h^{d-1}, and reads as follows

$$\frac{|\partial K|}{|K|} \lesssim \frac{1}{h}. \tag{4.10}$$

In Definition 4.1 we assumed the boundary of D to be $C^{1,1}$, i.e. C^1 with Lipschitz derivative. This requirement is sufficient because with such regularity ∂D satisfies the *uniform exterior ball condition*: For some $r_0 > 0$ and for all $x \in \partial\mathrm{D}$ it holds

$$\forall y \in \overline{\mathrm{D}} \setminus \{x\} : \frac{x-y}{|x-y|} \cdot n(x) + \frac{1}{2r_0}|x-y| \geq 0. \tag{4.11}$$

In order to define explicitly the numerical scheme that we are considering here, we first need to approximate the initial datum. Since the finite volume scheme approximates the solution by averaging on every cell, we can consider the discretization of the initial datum in this way,

$$\theta_K^0 = \fint_K \theta^0 \, \mathrm{d}x \tag{4.12}$$

and hence $\theta_h^0(x) = \theta_K^0$ for every $x \in K$ and every $K \in \mathcal{T}$. Since the scheme considers net fluxes across the cell faces, we define the discretized normal velocity from a control volume K to a neighboring one $L \sim K$ by

$$u_{KL}^n = \fint_{t^n}^{t^{n+1}} \fint_{K|L} u \cdot n_{KL} \, \mathrm{d}\mathcal{H}^{d-1} \, \mathrm{d}t. \tag{4.13}$$

Both u_{KL}^n and θ_K^0 are well-defined thanks to the trace theorem for Sobolev vector fields, i.e. (4.9). Notice that by definition the discretization of the velocity is antisymmetric with respect to the control volumes, i.e. it holds $u_{KL}^n = -u_{LK}^n$, which is useful for many calculations.

We define the finite volume scheme for the advection-diffusion equation (4.1) as

$$\frac{\theta_K^{n+1} - \theta_K^n}{k} + \sum_{L \sim K} \frac{|K|L|}{|K|} (u_{KL}^{n+}\theta_K^{n+1} - u_{KL}^{n-}\theta_L^{n+1}) + \kappa \sum_{L \sim K} \frac{|K|L|}{|K|} \frac{\theta_K^{n+1} - \theta_L^{n+1}}{d_{KL}} = 0 \tag{4.14}$$

for every $n \in [\![0, N-1]\!]$ and $K \in \mathcal{T}$. Therefore the *approximate solution* $\theta_{k,h}$ is defined by

$$\theta_{k,h}(t,x) = \theta_K^n \quad \text{for almost every } (t,x) \in [t^n, t^{n+1}) \times K$$

for every $n \in [\![0, N-1]\!]$ and $K \in \mathcal{T}$. If $n = 0$ we directly define $\theta_{k,h}^0 = \theta_h^0$.

Within the next section we show that this numerical problem is well-posed (see Lemma 4.1) and we will derive analogous stability estimate to (4.5) and (4.6) (see Lemma 4.2). These results follow under the assumption that the time step verifies $k \leq k_{\max}$. The definition of the *maximal time step* $k_{\max}$ follows a similar construction as in [16, 80], where it is given depending on some $\alpha > 1$ as the smaller number $k_{\max} = k_{\max}(\alpha)$ such that

$$\frac{q-1}{q}\int_I \|(\nabla\cdot u)^-\|_{L^\infty}\,\mathrm{d}t \le \frac{\alpha-1}{\alpha} \qquad \forall I\subseteq[0,T) \text{ with } |I|\le k_{\max}(\alpha). \tag{4.15}$$

The constant $\alpha > 1$ is used as a measure of how close the numerical solution $\theta_{k,h}$ is from satisfying the a priori estimate (4.5). Indeed, we will see in Lemma 4.2 that the exponent $1-1/q$ on the compressibility constant is replaced by $\alpha(1-1/q)$ and thus $\alpha = 1$ for incompressible vector fields, i.e. if $\nabla\cdot u = 0$.

4.2.2. Main result

The main result here presented concerns an estimate for the error generated by the finite volume scheme (4.14) as an approximation of the advection-diffusion equation (4.1). Without further ado let us recall the precise hypotheses we need for Theorem 4.1. First of all we consider a vector field in the DiPerna–Lions setting for some $p\in(1,\infty]$. Then assume the initial datum is integrable with

$$\theta^0\in L^q(\mathrm{D}) \qquad \text{with } q\in(1,\infty] \quad \text{such that} \quad \frac{1}{p}+\frac{1}{q}\le 1. \tag{4.16}$$

Last, for the numerical analysis we need to consider bounded vector fields,

$$u\in L^\infty((0,T)\times\mathrm{D}). \tag{4.17}$$

Although this is not required for the derivation of the continuous stability estimates (4.5)–(4.6), it is a standard and not very restrictive assumption for numerical experiments, see for instance [80]. The main result therefore states as follows.

Theorem 4.1. *Consider θ^0, u and $k_{\max}$ such that* (4.3)*,* (4.15)*,*(4.16) *and* (4.17) *hold. Consider an admissible tessellation of* D *that satisfies* (4.9)*. Let θ be the unique distributional solution to* (4.1)*–*(4.2) *and for $k\in(0,\min\{k_{\max},1\})$ and $h\in(0,1)$ let $\theta_{k,h}$ be the unique approximate solution given by the numerical scheme* (4.14)*. Then, for any $\delta>0$ there holds*

$$\sup_{0\le t\le T}\mathcal{D}_\delta(\theta(t),\theta_{k,h}(t)) \lesssim 1+\frac{h}{\delta}\min\left\{\|u\|_\infty\sqrt{\frac{T}{\kappa}},\sqrt{\frac{T\|u\|_\infty}{h}}\right\}+\frac{\sqrt{k}}{\delta}\left(\sqrt{T}\|u\|_\infty+\sqrt{\kappa}\right). \tag{4.18}$$

Here, $\mathcal{D}_\delta(\cdot,\cdot)$ defined as in (4.7), refers to a distance from the theory of optimal transportation with a logarithmic cost. This particular distance is of great use for equations with rough coefficients because it metrizes weak convergence. Namely, if we choose $\delta = h + \sqrt{k}$ then we get

$$\theta_{k,h} \rightharpoonup \theta \qquad \text{as } k, h \to 0$$

weakly with rate (at most) $\delta = h + \sqrt{k}$, we elaborate more on the properties of the Kantorovich–Rubinstein distance in Section 2.2.

The estimate (4.18) here is presented in such form to make explicit that it is a generalization of the result for transport equations ($\kappa = 0$) in [80]. One can appreciate that assuming $\kappa > 0$, Theorem 4.1 reads

$$\sup_{0\le t\le T} \mathcal{D}_\delta(\theta(t), \theta_{k,h}(t)) \lesssim 1 + \frac{h\|u\|_\infty}{\delta}\sqrt{\frac{T}{\kappa}} + \frac{\sqrt{k}}{\delta}\Big(\sqrt{T}\|u\|_\infty + \sqrt{\kappa}\Big),$$

which yields an error of order $\mathcal{O}(h + \sqrt{k})$ as stated before.

Another consequence of Theorem 4.1 is that in the limit $\kappa \to 0$ the estimate (4.18) takes the form

$$\sup_{0\le t\le T} \mathcal{D}_\delta(\theta(t), \theta_{k,h}(t)) \lesssim 1 + \frac{\sqrt{hT\|u\|_\infty}}{\delta} + \frac{\sqrt{kT}\|u\|_\infty}{\delta},$$

and thus recovers the result from [80], i.e. an error of order $\mathcal{O}(\sqrt{h} + \sqrt{k})$.

For the diffusionless transport equation, the main source of the spatial discretization error is the phenomenon of numerical diffusion. That is, the numerical scheme *acts like* creating a diffusion term with diffusion coefficient h, that for the transport equation reads as an advection-diffusion equation of the form

$$\partial_t\theta + \nabla\cdot(u\theta) = h\Delta\theta.$$

Therefore one would expect the rate of convergence for h to be of order $1/2$ since that is the known optimal rate for the vanishing diffusion or inviscid limit case, see [84]. However, with $\kappa > 0$, numerical diffusion acts modifying the already existing diffusion coefficient κ to $\kappa + h$

$$\partial_t\theta + \nabla\cdot(u\theta) = (\kappa + h)\Delta\theta,$$

and hence by the recent result [72], the expected rate of convergence for h has to be of order 1.

Let us close the discussion of the main result, by remarking on including source-sink distributions and non-homogeneous boundary conditions for (4.1). First, a flux boundary condition where (4.2) is replaced by $(\kappa\nabla\theta - u)\cdot n = g$ in $(0,T)\times\partial\mathrm{D}$ for some $g:[0,T]\times\partial\mathrm{D}\to\mathbb{R}$ can be transformed into a source-sink distribution $f:[0,T]\times\mathrm{D}\to\mathbb{R}$ in the domain using a suitable extension. Hence, it is sufficient to do the analysis instead of (4.1) for

$$\partial_t\theta + \nabla\cdot(u\theta) = \kappa\Delta\theta + f \qquad \text{in } (0,T)\times\mathrm{D}.$$

with suitable initial data and no-flux boundary condition (4.2). By another standard transformation, which consists of renormalizing the density θ, we can ensure that the total sources and sinks are balanced, which amounts to $\int_{\mathrm{D}} f(t,x)dx = 0$ for all $t\in[0,T]$. In particular, these transformations ensure that θ conserves mass. This is essential for using the optimal transport distance $\mathcal{D}_\delta$ to compare the solution θ with its numerical approximation $\theta_{k,h}$.

This situation with a balanced source-sink distribution was investigated in [80] for the diffusionless transport equation, and we expect that the analysis carries over to the present case with diffusion. The source-sink term will introduce additional discretization errors when we discuss the discretization of data in Lemmas 4.4–4.6 below. For the temporal discretization, where we use a stochastic Lagrangian representation of the solution, becomes more involved in the presence of a source-sink distribution and we omit it for the sake of concise presentation.

4.2.3. Properties of the numerical scheme

First of all we state a result on the well-posedness of the numerical scheme.

Lemma 4.1. *Under the hypothesis for Theorem 4.1, there exists a unique solution to the implicit upwind scheme* (4.21) *that is mass preserving and monotone, i.e. the solution remains positive for positive initial data.*

This is a classical result and we refer to [43, Theorem 4.1] for a detailed proof.

The main goal of this section is to develop stability estimates for the numerical scheme that are analogous to the a priori estimates (4.5) and (4.6). In order to do so it is convenient first to recall that some of the discretized versions of the functions involved on the scheme are controlled by their continuous counterpart. Specifically, recall that for the initial datum θ_h^0 and the divergence of the velocity field $\nabla\cdot u$ it holds

$$\|\theta_h^0\|_{L^q} \le \|\theta^0\|_{L^q}, \tag{4.19}$$

$$\|(\nabla\cdot u)_{k,h}^-\|_{L^1(L^\infty)} \le \|(\nabla\cdot u)^-\|_{L^1(L^\infty)}. \tag{4.20}$$

We omit the proof for the sake of brevity but it can be found on [80, Lemma 3].

Let us now rewrite the upwind scheme (4.14) in the following equivalent form:

$$\frac{\theta_K^{n+1}-\theta_K^n}{k}+\sum_{L\sim K}\frac{|K|L|}{|K|}u_{KL}^n\frac{\theta_K^{n+1}+\theta_L^{n+1}}{2}+\sum_{L\sim K}\frac{|K|L|}{|K|}|u_{KL}^n|\frac{\theta_K^{n+1}-\theta_L^{n+1}}{2} + \kappa\sum_{L\sim K}\frac{|K|L|}{|K|}\frac{\theta_K^{n+1}-\theta_L^{n+1}}{d_{KL}}=0 \tag{4.21}$$

for every $n\in[\![0,N-1]\!]$ and $K\in\mathcal{T}$. This is a straightforward consequence of the identities

$$u_{KL}^{n+}=\frac{|u_{KL}|+u_{KL}}{2}\qquad\text{and}\qquad u_{KL}^{n-}=\frac{|u_{KL}|-u_{KL}}{2}.$$

Then, the stability estimates for the finite volume scheme hold as follows.

Lemma 4.2 (Stability estimates)**.** *Let $\theta_{k,h}$ be the solution to the upwind scheme* (4.14) *with nonnegative initial data. Then for any $q\in(1,\infty)$, $\alpha>1$ and $k\le k_{\max}(\alpha)$ as defined in* (4.15)*, it holds*

$$\|\theta_{k,h}\|_{L^\infty(L^q)}\le\Lambda_{k,h}^{\alpha\left(1-\frac{1}{q}\right)}\|\theta_h^0\|_{L^q} \tag{4.22}$$

where $\Lambda_{k,h}=\exp(\|(\nabla\cdot u)_{k,h}^-\|_{L^1(L^\infty)})$. Moreover, if $r\in(1,\min\{q,2\}]$ it also holds,

$$\begin{aligned}&\sum_n\sum_K|K|\left(\frac{\theta_K^{n+1}+\theta_K^n}{2}\right)^{r-2}(\theta_K^{n+1}-\theta_K^n)^2\\&+k\sum_n\sum_K\sum_{L\sim K}|K|L|\left(|u_{KL}^n|+\frac{\kappa}{d_{KL}}\right)(\theta_K^{n+1}-\theta_L^{n+1})^2\left(\frac{\theta_K^{n+1}+\theta_L^{n+1}}{2}\right)^{r-2}\\&\le C_r\left(1+(r-1)\log\Lambda_{k,h}\right)\Lambda_{k,h}^{\alpha(r-1)}\|\theta_h^0\|_{L^r}^r\end{aligned} \tag{4.23}$$

with C_r being a positive constant that satisfies $C_r\to\infty$ as $r\to1$.

Proof. By the monotonicity of the scheme and the nonnegativity of the initial datum, we deduce that the solution of the numerical scheme $\theta_{k,h}$ is nonnegative. In order to study those stability estimates we will work with the second formulation of the upwind scheme (4.21). First of all, let us multiply the scheme by $|K|$ so that we get

$$|K|(\theta_K^{n+1}-\theta_K^n)+k\sum_{L\sim K}|K|L|u_{KL}^n\frac{\theta_K^{n+1}+\theta_L^{n+1}}{2} + k\sum_{L\sim K}|K|L||u_{KL}^n|\frac{\theta_K^{n+1}-\theta_L^{n+1}}{2}+\kappa k\sum_{L\sim K}|K|L|\frac{\theta_K^{n+1}-\theta_L^{n+1}}{d_{KL}}=0.$$

We denote the four addends as $\mathrm{I}_K^n+\mathrm{II}_K^n+\mathrm{III}_K^n+\mathrm{IV}_K^n=0$. Analogously to the continuous setting, we will obtain the stability estimates by testing with $(\theta_K^{n+1})^{q-1}$ and summing over $K \in \mathcal{T}$, namely

$$\underbrace{\sum_K \mathrm{I}_K^n(\theta_K^{n+1})^{q-1}}_{\mathrm{I}^n} + \underbrace{\sum_K \mathrm{II}_K^n(\theta_K^{n+1})^{q-1}}_{\mathrm{II}^n} + \underbrace{\sum_K \mathrm{III}_K^n(\theta_K^{n+1})^{q-1}}_{\mathrm{III}^n} + \underbrace{\sum_K \mathrm{IV}_K^n(\theta_K^{n+1})^{q-1}}_{\mathrm{IV}^n} = 0.$$

For the first term, we can apply Hölder's inequality,

$$\mathrm{I}^n = \sum_K |K|(\theta_K^{n+1})^q - \sum_K |K|\theta_K^n(\theta_K^{n+1})^{q-1} \geq \|\theta_{k,h}^{n+1}\|_{L^q}^q - \|\theta_{k,h}^n\|_{L^q}\|\theta_{k,h}^{n+1}\|_{L^q}^{q-1}.$$

For the second term we recall that $u_{KL}^n = -u_{LK}^n$, hence we can symmetrize II^n as follows

$$\mathrm{II}^n = \frac{k}{2}\sum_K \sum_{L\sim K} |K|L|u_{KL}^n \frac{\theta_K^{n+1} + \theta_L^{n+1}}{2}((\theta_K^{n+1})^{q-1} - (\theta_L^{n+1})^{q-1}).$$

We introduce the q-mean defined as a function $\Theta_q : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ such that

$$\Theta_q(x,y) = \frac{q-1}{q}\frac{x^q - y^q}{x^{q-1} - y^{q-1}}.$$

Note that $\Theta_2(x,y)$ is the arithmetic mean. Now, the above expression can be split into two factors, $\mathrm{II}^n = \mathrm{II}_1^n + \mathrm{II}_2^n$, defined as

$$\mathrm{II}_1^n = \frac{q-1}{q}\frac{k}{2}\sum_K \sum_{L\sim K} |K|L|u_{KL}^n((\theta_K^{n+1})^q - (\theta_L^{n+1})^q),$$

$$\mathrm{II}_2^n = \frac{k}{2}\sum_K \sum_{L\sim K} |K|L|(\Theta_2 - \Theta_q)(\theta_K^{n+1}, \theta_L^{n+1})((\theta_K^{n+1})^{q-1} - (\theta_L^{n+1})^{q-1}).$$

On the one hand, for the first addend we can symmetrize again such that

$$\begin{aligned}\mathrm{II}_1^n &= \frac{q-1}{q}k\sum_K(\theta_K^{n+1})^q \sum_{K\sim L}|K|L|u_{KL}^n = \frac{q-1}{q}k\sum_K(\theta_K^{n+1})^q(\nabla\cdot u)_K^n \\ &\geq -\frac{q-1}{q}\lambda^n\|\theta_{k,h}(t^{n+1})\|_{L^q}^q\end{aligned}$$

where $\lambda^n = k\|(\nabla\cdot u(t^n))_{k,h}^-\|_{L^\infty}$. On the other hand, we estimate II_2^n using the following bound

$$|\Theta_2(x,y) - \Theta_q(x,y)| \leq \frac{|q-2|}{q}\frac{|x-y|}{2} \tag{4.24}$$

for all $x, y > 0$. More information about the q-mean and a detailed proof of the latter estimate can be found on [80, Appendix A]. By the estimate (4.24) follows

$$\mathrm{II}_2^n \geq -\frac{k}{2}\frac{|q-2|}{q}\sum_K\sum_{L\sim K}|K|L||u_{KL}^n|\frac{\theta_K^{n+1}-\theta_L^{n+1}}{2}((\theta_K^{n+1})^{q-1}-(\theta_L^{n+1})^{q-1}).$$

Analogously, for both III^n and IV^n the symmetrization procedure might be applied to get the bounds

$$\mathrm{III}^n \geq \frac{k}{2}\sum_K\sum_{L\sim K}|K|L||u_{KL}^n|\frac{\theta_K^{n+1}-\theta_L^{n+1}}{2}((\theta_K^{n+1})^{q-1}-(\theta_L^{n+1})^{q-1}).$$

$$\mathrm{IV}^n = \kappa\frac{k}{2}\sum_K\sum_{L\sim K}|K|L|\frac{\theta_K^{n+1}-\theta_L^{n+1}}{d_{KL}}((\theta_K^{n+1})^{q-1}-(\theta_L^{n+1})^{q-1}).$$

All in all we get the estimate

$$\begin{aligned}\|\theta_{k,h}^{n+1}\|_{L^q}^q &+ \frac{k}{2}\left(1-\frac{|q-2|}{q}\right)\sum_K\sum_{L\sim K}|K|L||u_{KL}^n|\frac{\theta_K^{n+1}-\theta_L^{n+1}}{2}((\theta_K^{n+1})^{q-1}-(\theta_L^{n+1})^{q-1})\\ &+\kappa\frac{k}{2}\sum_K\sum_{L\sim K}|K|L|\frac{\theta_K^{n+1}-\theta_L^{n+1}}{d_{KL}}((\theta_K^{n+1})^{q-1}-(\theta_L^{n+1})^{q-1})\\ &\leq \|\theta_{k,h}^n\|_{L^q}\|\theta_{k,h}^{n+1}\|_{L^q}^{q-1}+\frac{q-1}{q}\lambda^n\|\theta_{k,h}^{n+1}\|_{L^q}^q.\end{aligned} \tag{4.25}$$

In order to obtain the first stability estimate (4.22) we can drop the second and third addends in (4.25) such that, dividing by $\|\theta_{k,h}^{n+1}\|_{L^q}^{q-1}$, we get

$$\left(1-\frac{q-1}{q}\lambda^n\right)\|\theta_{k,h}^{n+1}\|_{L^q}\leq\|\theta_{k,h}^n\|_{L^q}.$$

Now, if $k\leq k_{\max}(\alpha)$ it holds that

$$\frac{q-1}{q}\lambda^n\leq\frac{\alpha-1}{\alpha},$$

and therefore

$$\frac{1}{1-\frac{q-1}{q}\lambda^n}\leq 1+\alpha\frac{q-1}{q}\lambda^n\leq\exp\left(\alpha\frac{q-1}{q}\lambda^n\right).$$

By an iterative argument we get

$$\|\theta_{k,h}^n\|_{L^q} \le \exp\left(\alpha\frac{q-1}{q}k\sum_{i=1}^{n}\|(\nabla\cdot u(t^{i-1}))_{k,h}^-\|_{L^\infty}\right)\|\theta_h^0\|_{L^q}$$

for every $n \in [\![0, N]\!]$ and thus we get the first stability estimate (4.22).

To establish the temporal and spatial gradient estimate (4.23) we repeat a similar computation. However now we need to develop a different bound for the term I^n and thus we use the estimate

$$rx^{r-1}(x-y) \ge x^r - y^r + \frac{r(r-1)}{2^{3-r}}\left(\frac{x+y}{2}\right)^{r-2}(x-y)^2$$

that holds for $r \in (1,2]$ and comes from the convexity of the map $x \mapsto x^r$. Then, by setting $x = \theta_K^{n+1}$, $y = \theta_K^n$ and $r \in (1, \min\{q,2\}]$, we get the following lower bound for I^n,

$$\begin{aligned}\mathrm{I}^n &= \sum_K |K|(\theta_K^{n+1})^r - \sum_K |K|\theta_K^n(\theta_K^{n+1})^{r-1} = \sum_K |K|(\theta_K^{n+1})^{r-1}(\theta_K^{n+1}-\theta_K^n)\\ &\ge \frac{1}{r}\sum_K |K|((\theta_K^{n+1})^r - (\theta_K^n)^r) + \frac{r-1}{2^{3-r}}\sum_K |K|\left(\frac{\theta_K^{n+1}+\theta_K^n}{2}\right)^{r-2}(\theta_K^{n+1}-\theta_K^n)^2,\end{aligned}$$

and adding it to the stability estimate (4.25) instead of the previous one, we get

$$\begin{aligned}&\frac{1}{r}\|\theta_{k,h}^{n+1}\|_{L^r}^r + \frac{r-1}{2^{3-r}}\sum_K |K|\left(\frac{\theta_K^{n+1}+\theta_K^n}{2}\right)^{r-2}(\theta_K^{n+1}-\theta_K^n)^2\\ &\qquad + \frac{r-1}{r}\frac{k}{2}\sum_K\sum_{L\sim K}|K|L||u_{KL}^n|(\theta_K^{n+1}-\theta_L^{n+1})((\theta_K^{n+1})^{r-1}-(\theta_L^{n+1})^{r-1})\\ &\qquad + \kappa\frac{k}{2}\sum_K\sum_{L\sim K}|K|L|\frac{\theta_K^{n+1}-\theta_L^{n+1}}{d_{KL}}((\theta_K^{n+1})^{r-1}-(\theta_L^{n+1})^{r-1})\\ &\qquad \le \frac{1}{r}\|\theta_{k,h}^n\|_{L^r}^r + \frac{r-1}{r}\lambda^n\|\theta_{k,h}^{n+1}\|_{L^r}^r.\end{aligned}$$

We now rewrite the advection and diffusion terms using the following elementary inequality that holds for any $r \in (1,2]$ and $x, y > 0$,

$$(x-y)^2\left(\frac{x+y}{2}\right)^{r-2} \le (x-y)\frac{x^{r-1}-y^{r-1}}{r-1}.$$

Choosing $x = \theta_K^{n+1}$, $y = \theta_L^{n+1}$ we thus get

$$\begin{aligned}
&\frac{r-1}{2^{3-r}}\sum_K |K| \left(\frac{\theta_K^{n+1}+\theta_K^n}{2}\right)^{r-2} (\theta_K^{n+1}-\theta_K^n)^2 \\
&\quad + \frac{r-1}{r}\frac{k}{2}\sum_K\sum_{L\sim K} |K|L| \left(|u_{KL}^n| + \frac{\kappa}{d_{KL}}\right)\left(\frac{\theta_K^{n+1}-\theta_L^{n+1}}{2}\right)^{r-2} (\theta_K^{n+1}-\theta_L^{n+1})^2 \\
&\quad \le \frac{1}{r}\|\theta_{k,h}^n\|_{L^r}^r + \frac{r-1}{r}\lambda^n\|\theta_{k,h}^{n+1}\|_{L^r}^r.
\end{aligned}$$

Summing over n and applying (4.22) it yields the desired estimate (4.23) with constant

$$C_r = 2\frac{\max\{2^{2-r}, r\}}{r(r-1)}. \qquad \blacksquare$$

Lemma 4.2 provides a discrete version of the standard stability and energy estimates in the continuous setting. On the one hand (4.22) is the discrete version of (4.5), while on the other hand (4.6) is reproduced in the numerical scheme setting by (4.23) dropping the addends related to the time derivative and the advection field.

A direct consequence of Lemma 4.2 together with (4.19) and (4.20) is that the expressions on the right hand side in (4.22) and (4.23) are controlled by $\|\theta^0\|_{L^q}$ and $\|(\nabla\cdot u)^-\|_{L^1(L^\infty)}$ and therefore they are $\mathcal{O}(1)$. In particular it holds

$$\|\theta_{k,h}\|_{L^\infty(L^q)} \lesssim 1,$$

which is certainly not surprising since that also holds for the exact solutions of (4.1).

Let us introduce now two weak BV estimates which will be a key tool to obtain the desired result from Theorem 4.1. These estimates are a consequence of numerical diffusion.

Lemma 4.3 (BV estimates)**.** *Let $\theta_{k,h}$ be a solution of the numerical scheme* (4.14)*. Under the assumptions of Theorem 4.1 we get the following BV estimates*

$$\sum_n\sum_K |K||\theta_K^{n+1}-\theta_K^n| \lesssim \sqrt{\frac{T}{k}}, \tag{4.26}$$

$$k\sum_n\sum_K\sum_{L\sim K} |K|L|\left(\sqrt{\frac{\kappa}{T}} + \sqrt{\frac{h}{T\|u\|_\infty}}|u_{KL}^n|\right)|\theta_K^{n+1}-\theta_L^{n+1}| \lesssim 1. \tag{4.27}$$

The first estimate on the time discretization (4.26) does not have a counterpart in the continuous setting and it is a by-product of the numerical diffusion introduced by the temporal discretization of the scheme. The second one (4.27) instead presents two differentiated parts. First we obtain a spatial *strong* BV estimate

$$k\sum_n\sum_K\sum_{L\sim K}|K|L||\theta_K^{n+1}-\theta_L^{n+1}|\lesssim\sqrt{\frac{T}{\kappa}},\tag{4.28}$$

which is precisely the responsible for carrying an upgrade on the convergence rate from $\mathcal{O}(h^{1/2})$ to $\mathcal{O}(h)$ in comparison with the transport equation without diffusion, [79, 80]. This BV estimate can be understood as the discrete analogous to

$$\|\nabla\theta\|_{L^1(L^1)}\lesssim\sqrt{\frac{T}{\kappa}}.$$

Then we also obtain the spatial *weak* BV estimate

$$k\sum_n\sum_K\sum_{L\sim K}|K|L||u_{KL}^n||\theta_K^{n+1}-\theta_L^{n+1}|\lesssim\sqrt{\frac{T\|u\|_\infty}{h}}\tag{4.29}$$

that is a consequence of the numerical diffusion introduced by the spatial discretization and can be read as the surviving part in the limit $\kappa\to 0$. It is precisely the weak BV estimate obtained in [80, Proposition 1] for the transport equation.

Proof. We start proving (4.26). Let us first consider a nonnegative initial datum. Let $r\in(1,\min\{2,q\}]$ and smuggle into (4.26) the weight $((\theta_K^{n+1}+\theta_K^n)/2)^{(r-2)/2}$ such that

$$\sum_K|K||\theta_K^{n+1}-\theta_K^n|\left(\frac{\theta_K^{n+1}+\theta_K^n}{2}\right)^{\frac{r-2}{2}}\left(\frac{\theta_K^{n+1}+\theta_K^n}{2}\right)^{\frac{2-r}{2}}=\mathrm{I}^n\,.$$

Then, via Cauchy-Schwarz,

$$\mathrm{I}^n\le\left[\sum_K|K|(\theta_K^{n+1}-\theta_K^n)^2\left(\frac{\theta_K^{n+1}+\theta_K^n}{2}\right)^{r-2}\right]^{1/2}\left[\sum_K|K|\left(\frac{\theta_K^{n+1}+\theta_K^n}{2}\right)^{2-r}\right]^{1/2}.$$

By Lemma 4.2 and (4.20), the first factor of the product is controlled by a constant depending on r, the $L^1(L^\infty)$ norm of $(\nabla\cdot u)^-$ and the L^r norm of the initial datum. Therefore, summing over n and applying Jensen's inequality for the time variable now we can write,

$$\sum_n \mathrm{I}^n \lesssim \sum_n \left[\sum_K |K| \left(\frac{\theta_K^{n+1}+\theta_K^n}{2}\right)^{2-r}\right]^{1/2} \leq \sum_n \left[\sum_K |K|((\theta_K^{n+1})^{2-r} + (\theta_K^n)^{2-r})\right]^{1/2}$$
$$\leq 2T^{1/2}\left(\sum_n \|\theta_{k,h}(t^n)\|_{L^{2-r}}^{2-r}\right)^{1/2} \lesssim \sqrt{\frac{T}{k}}\|\theta_{k,h}\|_{L^1(L^{2-r})}^{(2-r)/2}.$$

Hence, by (4.19) and (4.20) we get the weak BV estimate (4.26) for nonnegative initial data. Once this is established, for general initial data the estimate follows via triangle inequality.

We argue analogously to get the estimate (4.27). We will obtain the two estimates (4.28) and (4.29) separately. Let us start with the strong BV estimate (4.28). Consider a non negative initial datum since for a general case we can just apply a triangle inequality. Smuggling the same weight as before, with $r \in (1, \min\{q, 2\}]$, together with a factor d_{KL} we can write via Cauchy-Schwarz inequality,

$$k\sum_n \sum_K \sum_{L\sim K} |K|L||\theta_K^{n+1} - \theta_L^{n+1}| = k\sum_n (\mathrm{II}_S^n)^{1/2}(\mathrm{III}_S^n)^{1/2} = \left(k\sum_n \mathrm{II}_S^n\right)^{1/2}\left(k\sum_n \mathrm{III}_S^n\right)^{1/2}$$

with

$$\mathrm{II}_S^n = \sum_K \sum_{L\sim K} |K|L| \frac{(\theta_K^{n+1} - \theta_L^{n+1})^2}{d_{KL}} \left(\frac{\theta_K^{n+1}+\theta_L^{n+1}}{2}\right)^{r-2},$$
$$\mathrm{III}_S^n = \sum_K \sum_{L\sim K} |K|L| d_{KL} \left(\frac{\theta_K^{n+1}+\theta_L^{n+1}}{2}\right)^{2-r}.$$

The term involving II_S^n is controlled thanks to (4.23) by

$$\left(k\sum_n \mathrm{II}_S^n\right)^{1/2} \lesssim \frac{1}{\sqrt{\kappa}}.$$

For III_S^n we can use the identity $((x+y)/2)^{2-r} \leq x^{2-r} + y^{2-r}$ for any $x, y > 0$ and the trivial bound $d_{KL} \leq 2h$. Then by the isoperimetric property of the mesh (4.10) we get

$$\mathrm{III}_S^n \leq h\sum_K (\theta_K^{n+1})^{2-r} \sum_{L\sim K} |K|L| \lesssim \sum_K |K|(\theta_K^{n+1})^{2-r} = \|\theta_{k,h}(t^n)\|_{L^{2-r}}.$$

Again we can estimate $\|\theta_{k,h}(t^n)\|_{L^{2-r}}$ by $\|\theta_{k,h}(t^n)\|_{L^r}$ and a factor depending on |D| so that it yields the remaining term

$$\left(k\sum_n \mathrm{III}_S^n\right)^{1/2} \lesssim \sqrt{T}.$$

Thus, we obtain the strong BV estimate (4.28).

For the weak BV estimate (4.29) we follow a similar argument. Let $r \in (1, \min\{q, 2\}]$ and apply Cauchy-Schwarz as before,

$$\begin{aligned} k\sum_n\sum_K\sum_{L\sim K} |K|L||u_{KL}^n||\theta_K^{n+1} - \theta_L^{n+1}| &= k\sum_n (\mathrm{II}_W^n)^{1/2}(\mathrm{III}_W^n)^{1/2} \\ &= \left(k\sum_n \mathrm{II}_W^n\right)^{1/2}\left(k\sum_n \mathrm{III}_W^n\right)^{1/2} \end{aligned}$$

where now we define

$$\begin{aligned} \mathrm{II}_W^n &= \sum_K\sum_{L\sim K} |K|L||u_{KL}^n|^2(\theta_K^{n+1} - \theta_L^{n+1})^2 \left(\frac{\theta_K^{n+1} + \theta_L^{n+1}}{2}\right)^{r-2}, \\ \mathrm{III}_W^n &= \sum_K\sum_{L\sim K} |K|L| \left(\frac{\theta_K^{n+1} + \theta_L^{n+1}}{2}\right)^{2-r}. \end{aligned}$$

Then a direct application of (4.23) and following the previous argument for the strong BV estimate we obtain

$$\left(k\sum_n \mathrm{II}_W^n\right)^{1/2} \lesssim \sqrt{\|u\|_\infty} \qquad \text{and} \qquad \left(k\sum_n \mathrm{III}_W^n\right)^{1/2} \lesssim \sqrt{\frac{T}{h}}$$

so that we complete the proof of (4.27). ■

4.3. Proof of Theorem 4.1

In this section we will prove the main result of the chapter. In order to do so we need to derive all the error estimates coming from the different discretizations that contribute to the stability estimate (4.18). There are two main sources of error: on the one hand the discretization in time and space of the initial datum and the vector field and on the other hand there is the error associated to the scheme, also known as truncation error. For the diffusionless transport equation one can see (for instance, in [79, 80]) that the error that governs the convergence of the numerical solution comes exclusively in form of truncation error. However in our case we will see how both sources of error, truncation and discretization of data, contribute equally to the final estimate.

Before turning to the proof of the Theorem let us first mention essential mathematical tools to study stability estimates for the advection-diffusion equations in a low regularity framework.

On the one hand, we use the Hardy-Littlewood maximal function from the Calderón-Zygmund theory in harmonic analysis. Given a measurable function $f : \mathbb{R}^d \to \mathbb{R}$, we say M is the maximal function operator and it is defined by

$$Mf(x) = \sup_{R>0} \frac{1}{R^d} \int_{B_R(x)\cap \mathrm{D}} |f(y)| \,\mathrm{d}y.$$

The operator is continuous from L^p to L^p for every $1 < p \leq \infty$ and therefore we get the estimate,

$$\|Mf\|_{L^p} \lesssim \|f\|_{L^p}, \quad \text{for } 1 < p \leq \infty. \tag{4.30}$$

Moreover, via the maximal function we can establish bounds for the different quotients of a measurable function through the so-called *Morrey's inequality*, that is

$$\frac{|f(x) - f(y)|}{|x-y|} \lesssim (M\nabla\overline{f})(x) + (M\nabla\overline{f})(y) \tag{4.31}$$

for almost every $x, y \in \mathrm{D}$ and where $\overline{f}$ denotes a Sobolev regular extension of f to the full space $\mathbb{R}^d$.

On the other hand, as stated in the introduction the stability estimate (4.6) provides for any $I \subset [0,T]$ an explicit control on the $L^1(I; W^{1,1}(\mathrm{D}))$ norm of the solution to (4.1). We can see this by choosing $r \in (1, \min\{q, 2\}]$, then we have

$$\begin{aligned} \int_I \int_{\mathrm{D}} |\nabla\theta| \,\mathrm{d}x\,\mathrm{d}t &\leq \left(\int_I \int_{\mathrm{D}} |\theta|^{r-2} |\nabla\theta|^2 \,\mathrm{d}x\,\mathrm{d}t\right)^{1/2} \left(\int_I \int_{\mathrm{D}} |\theta|^{2-r} \,\mathrm{d}x\,\mathrm{d}t\right)^{1/2} \\ &\lesssim \sqrt{\frac{|I|}{\kappa}} \|\theta^0\|_{L^r} \end{aligned} \tag{4.32}$$

where we have used Hölder's inequality and we have estimated $\|\theta\|_{L^\infty(L^{2-r})}$ by $\|\theta\|_{L^\infty(L^r)}$ with a factor depending on $|\mathrm{D}|$ and

$$\int_I \int_{\mathrm{D}} |\theta|^{2-r} \,\mathrm{d}x\,\mathrm{d}t \lesssim \int_I \left(\int_{\mathrm{D}} |\theta|^r \,\mathrm{d}x\right)^{\frac{2-r}{r}} \mathrm{d}t \lesssim |I| \|\theta\|^{2-r}_{L^\infty(L^r)} \lesssim |I| \|\theta^0\|^{2-r}_{L^r}.$$

4.3.1. Error due to the discretization of the data

We start with the contribution to the error estimates caused by the discretization in time.

Lemma 4.4. *Let $t \in [t^n, t^{n+1})$ with $n \in [\![0, N-1]\!]$. Then it holds*

$$\mathcal{D}_\delta(\theta(t), \theta(t^n)) \lesssim \frac{k\|u\|_\infty + \sqrt{k\kappa}}{\delta}. \tag{4.33}$$

Proof. Let ζ_t be the optimal Kantorovich potential corresponding to the distance $\mathcal{D}_\delta(\theta(t), \theta(t^n))$ at time $t \in [t^n, t^{n+1})$ for some $n \in [\![0, N-1]\!]$, such that

$$\mathcal{D}_\delta(\theta(t), \theta(t^n)) = \int_{\mathrm{D}} \zeta_t(x)(\theta(t,x) - \theta(t^n, x))\,\mathrm{d}x.$$

By means of (2.11) we can rewrite the distance as

$$\mathcal{D}_\delta(\theta(t), \theta(t^n)) = \int_{t^n}^{t} \int_{\mathrm{D}} \nabla\zeta_t(x) \cdot u(s,x)\theta(s,x)\,\mathrm{d}x\,\mathrm{d}s - \kappa \int_{t^n}^{t} \int_{\mathrm{D}} \nabla\zeta_t(x) \cdot \nabla\theta(s,x)\,\mathrm{d}x\,\mathrm{d}s,$$

that to shorten the notation we denote as $\mathcal{D}_\delta(\theta(t), \theta(t^n)) = \mathrm{I} + \mathrm{II}$. The first addend can be controlled with the properties of the Kantorovich potential as follows,

$$\mathrm{I} = \int_{t^n}^{t} \int_{\mathrm{D}} \nabla\zeta_t(x) \cdot u(s,x)\theta(s,x)\,\mathrm{d}x\,\mathrm{d}s \lesssim \frac{k}{\delta}\|u\|_\infty \|\theta\|_{L^\infty(L^1)}.$$

For the second term, we apply the estimate (4.32) on the time interval $[t^n, t)$ and we get

$$\|\nabla\theta\|_{L^1([t^n,t);L^1(\mathrm{D}))} \lesssim \sqrt{\frac{t-t^n}{\kappa}} \leq \sqrt{\frac{k}{\kappa}}$$

and thus it yields the bound for II via

$$\mathrm{II} = -\kappa \int_{t^n}^{t} \int_{\mathrm{D}} \nabla\zeta_t(x) \cdot \nabla\theta(s,x)\,\mathrm{d}x\,\mathrm{d}s \leq \frac{\kappa}{\delta} \int_{t^n}^{t} \int_{\mathrm{D}} |\nabla\theta(s,x)|\,\mathrm{d}x\,\mathrm{d}s \lesssim \frac{\sqrt{k\kappa}}{\delta}.$$

Thus, putting everything together it yields the estimate (4.33). ■

Next in order we study the error caused by the spatial discretization of the initial datum θ^0. We define $\theta_h^0(x) = \theta_K^0(x)$ as in (4.12) piecewise for almost every $x \in K$ and for each $K \in \mathcal{T}$. This result is a straightforward consequence of the stability estimate for the advection-diffusion equation (4.8).

Lemma 4.5. *Let θ^h be the solution to the advection-diffusion equation* (4.1) *with initial datum θ_h^0. Then it holds*

$$\sup_{0\le t\le T} \mathcal{D}_\delta(\theta(t),\theta^h(t)) \lesssim 1+\frac{h}{\delta}. \tag{4.34}$$

Proof. In this case θ and θ^h are solutions to the same equation with same velocity fields and same diffusion coefficients, therefore a direct application of (4.8) yields

$$\sup_{0\le t\le T} \mathcal{D}_\delta(\theta(t),\theta^h(t)) \lesssim 1+\mathcal{D}_\delta(\theta^0,\theta_h^0).$$

Now let us write ζ_t to denote the optimal Kantorovich potential such that it holds

$$\mathcal{D}_\delta(\theta^0,\theta_h^0) = \int_{\mathrm{D}} \zeta_t(x)(\theta^0(x)-\theta_h^0)\,\mathrm{d}x = \int_{\mathrm{D}} (\zeta_t(x)-(\zeta_t)_h(x))\theta^0(x)\,\mathrm{d}x$$

where the second equality comes from the symmetry property of the cell-averaging $(\cdot)_h$ operator, that is

$$\int_{\mathrm{D}} f(x)g_h(x)\,\mathrm{d}x = \sum_K |K| \fint_K \fint_K f(x)g(y)\,\mathrm{d}y\,\mathrm{d}x = \int_{\mathrm{D}} f_h(x)g(x)\,\mathrm{d}x$$

for all integrable f and g such that its product is also integrable. Furthermore, we use the definition of the Kantorovich potential together with its Lipschitz bound pointwise in $x\in K$ so that,

$$|\zeta_t(x)-(\zeta_t)_h(x)| \le \fint_K |\zeta_t(x)-\zeta_t(y)|\,\mathrm{d}y \le \fint_K \log\Big(\frac{|x-y|}{\delta}+1\Big)\,\mathrm{d}y \le \log\Big(\frac{h}{\delta}+1\Big) \le \frac{h}{\delta}.$$

We thus find the final estimate (4.34) just by combining everything. ■

In addition we must also consider the error due to the time discretization for the coefficients of the equation. We denote by u^k the vector field averaged in time over $[t^n,t^{n+1})$ as follows,

$$u^k(t,x) = \fint_{t^n}^{t^{n+1}} u(t,x)\,\mathrm{d}t \quad \text{for a.e. } t\in[t^n,t^{n+1}).$$

Lemma 4.6. *Let θ^k be the solution to the advection-diffusion equation* (4.1) *with vector field u^k. Then it holds for any $m\in[\![0,N]\!]$*

$$\mathcal{D}_\delta(\theta(t^m,\cdot),\theta^k(t^m,\cdot)) \lesssim 1+\frac{k(\|u\|_\infty+1)}{\delta}+\frac{\sqrt{k\kappa}}{\delta}. \tag{4.35}$$

For the proof of the Lemma we need to introduce a stochastic Lagrangian representation for the advection-diffusion equation. Consider a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$, for any $t \geq 0$ we say the map $X_t : \mathrm{D} \to \mathrm{D}$ is an *stochastic Lagrangian flow* if for every $x \in \mathrm{D}$ it solves the stochastic differential equation

$$X_t = X_t^x = x + \int_0^t u(s, X_s(x))\,\mathrm{d}s + \sqrt{2\kappa}\,B_t - \int_0^t n(X_s(x))\,\mathrm{d}L_s. \tag{4.36}$$

Here $\{B_t\}_{t\geq 0}$ is a $\mathcal{F}_t$-adapted Brownian motion and $\{L_t\}_{t\geq 0}$ is an $\mathcal{F}_t$-adapted local time of the process $\{X_t\}_{t\geq 0}$ at the boundary $\partial\mathrm{D}$. By the classic Doob maximal martingale inequality (see [77]), we have for any $q > 1$ the bound for the Brownian motion,

$$\mathbb{E}\left[\sup_{0\leq s\leq t} |B_s|^q\right]^{\frac{1}{q}} \leq \frac{q}{q-1}\sqrt{t}. \tag{4.37}$$

Since the setting and tools needed for the proof of this Lemma use some language from stochastic analysis and differs from the rest of the mathematical tools presented in this chapter, we include for the convenience of the reader the Appendix A.1 reviewing some of the abstract setting and the formal definitions that will be used along this proof.

Proof of Lemma 4.6. We can assume by a density argument that u and u^k are smooth in space and continuous in time. Indeed, this a consequence of a classic approximation argument leading to the emergence of a commutator, which can be estimated along the lines of [37, Lemma 2.1] or [33, Section 2], and the fact that the logarithmic Kantorovich–Rubinstein distance metrizes weak convergence.

Without loss of generality, we assume that θ^0 is a probability measure. Hence, by the results stated in the Appendix we find processes $\{X_t\}_{t\geq 0}$ and $\{X_t^k\}_{t\geq 0}$, strong solutions to the reflected SDE (A.3) started with law θ^0 driven by the same Brownian motion $\{B_t\}_{t\geq 0}$ with vector field u and u^k, respectively. The according local times at the boundary are denoted by $\{L_t\}_{t\geq 0}$ and $\{L_t^k\}_{t\geq 0}$. In this way, we constructed a pathwise coupling of $\theta(t)$ and $\theta^k(t)$, i.e. $\operatorname{law} X_t = \theta(t)$ and $\operatorname{law} X_t^k = \theta^k(t)$ and we can straightforwardly estimate the logarithmic Kantorovich–Rubinstein distance with the help of the Lagrangian coupling for any $t \in [0,T]$ by

$$\begin{aligned}
\mathcal{D}_\delta(\theta(t),\theta^k(t)) &\le \mathbb{E}_{\theta^0}\left[\log\left(\frac{|X_t - X_t^k|}{\delta} + 1\right)\right] \\
&\le e^{\frac{t}{r_0}}\mathbb{E}_{\theta^0}\left[\log\left(\frac{|X_t - X_t^k|}{\delta} e^{-\frac{1}{2r_0}\left(L_t + L_t^k\right)} + 1\right)\right], \\
&\lesssim \mathbb{E}_{\theta^0}\left[\log\left(\frac{|X_t - X_t^k|}{\delta} e^{-\frac{1}{2r_0}\left(L_t + L_t^k\right)} + 1\right)\right],
\end{aligned}$$

where we used the fact that the boundary local times satisfy $|L_t|, |L_t^k| \le t$ for any $t \in [0,T]$ and where r_0 is the constant given by the uniform exterior ball condition for the domain (4.11). We have also estimated e^{t/r_0} with e^{T/r_0} and absorbed this constant in $\lesssim$. Hence, by telescoping and using that $X_0 = X_0^k$, we arrive at the estimate

$$\begin{aligned}
\mathcal{D}_\delta(\theta(t^m),\theta^k(t^m)) \lesssim \sum_{n=0}^{m-1}\Bigg(&\mathbb{E}_{\theta^0}\left[\log\left(\frac{|X_{t^{n+1}} - X_{t^{n+1}}^k|}{\delta} e^{-\frac{1}{2r_0}\left(L_{t^{n+1}} + L_{t^{n+1}}^k\right)} + 1\right)\right] \\
&- \mathbb{E}_{\theta^0}\left[\log\left(\frac{|X_{t^n} - X_{t^n}^k|}{\delta} e^{-\frac{1}{2r_0}\left(L_{t^n} + L_{t^n}^k\right)} + 1\right)\right]\Bigg).
\end{aligned}$$

The representation (4.36) and Itô's formula allows to estimate for any $n \in [\![0, m-1]\!]$

$$\begin{aligned}
&\mathbb{E}_{\theta^0}\left[\log\left(\frac{|X_{t^{n+1}} - X_{t^{n+1}}^k|}{\delta} e^{-\frac{1}{2r_0}\left(L_{t^{n+1}} + L_{t^{n+1}}^k\right)} + 1\right)\right. \\
&\quad \left. - \log\left(\frac{|X_{t^n} - X_{t^n}^k|}{\delta} e^{-\frac{1}{2r_0}\left(L_{t^n} + L_{t^n}^k\right)} + 1\right)\right] \\
&= \mathbb{E}_{\theta^0}\left[\int_{t^n}^{t^{n+1}} e^{-\frac{1}{2r_0}\left(L_t + L_t^k\right)} \frac{\frac{X_t - X_t^k}{|X_t - X_t^k|}\cdot\left(\mathrm{d}X_t - \mathrm{d}X_t^k\right) - \frac{1}{2r_0}|X_t - X_t^k|\left(\mathrm{d}L_t + \mathrm{d}L_t^k\right)}{|X_t - X_t^k| e^{-\frac{1}{2r_0}\left(L_t + L_t^k\right)} + \delta}\right] \\
&\lesssim \mathbb{E}_{\theta^0}\left[\int_{t^n}^{t^{n+1}} \frac{\frac{X_t - X_t^k}{|X_t - X_t^k|}\cdot\left(\mathrm{d}X_t - \mathrm{d}X_t^k\right) - \frac{1}{2r_0}|X_t - X_t^k|\left(\mathrm{d}L_t + \mathrm{d}L_t^k\right)}{|X_t - X_t^k| e^{-\frac{1}{2r_0}\left(L_t + L_t^k\right)} + \delta}\right] \\
&\le \mathbb{E}_{\theta^0}\left[\int_{t^n}^{t^{n+1}} \frac{\left|u(t,X_t) - u^k(t,X_t^k)\right|}{|X_t - X_t^k| e^{-\frac{1}{2r_0}\left(L_t + L_t^k\right)} + \delta}\, dt\right] \\
&\quad - \mathbb{E}_{\theta^0}\left[\int_{t^n}^{t^{n+1}} \frac{\frac{X_t - X_t^k}{|X_t - X_t^k|}\cdot n(X_t)\, dL_t - \frac{X_t - X_t^k}{|X_t - X_t^k|}\cdot n(X_t^k)\, dL_t^k + \frac{1}{2r_0}|X_t - X_t^k|(\mathrm{d}L_t + \mathrm{d}L_t^k)}{|X_t - X_t^k| e^{-\frac{1}{2r_0}\left(L_t + L_t^k\right)} + \delta}\right].
\end{aligned}$$

Next, we rearrange the integrands in the dominator of the second term in such a way that those have a sign thanks to the exterior ball condition (4.11). Indeed, we observe that for $X_t, X_t^k \in \overline{\mathrm{D}}$, one has

$$\frac{X_t - X_t^k}{|X_t - X_t^k|} \cdot n(X_t) + \frac{1}{2r_0}\Big|X_t - X_t^k\Big| \geq 0,$$

and

$$-\frac{X_t - X_t^k}{|X_t - X_t^k|} \cdot n(X_t^k) + \frac{1}{2r_0}\Big|X_t - X_t^k\Big| = \frac{X_t^k - X_t}{|X_t^k - X_t|} \cdot n(X_t^k) + \frac{1}{2r_0}\Big|X_t^k - X_t\Big| \geq 0.$$

Hence, it is enough to continue to estimate the first one, for which we first get rid of the exponential factor in the denominator again using the property $|L_t|, |L_t^k| \leq t$. Summarizing our findings so far, we get

$$\mathcal{D}_\delta(\theta(t^m), \theta^k(t^m)) \lesssim \exp\Big(\frac{2t^m}{r_0}\Big) \sum_{n=0}^{m-1} \mathrm{I}^n \tag{4.38}$$

where

$$\mathrm{I}^n = \mathbb{E}_{\theta^0}\left[\int_{t^n}^{t^{n+1}} \frac{\Big|u(t, X_t) - u^k(s, X_t^k)\Big| \,\mathrm{d}s}{|X_t - X_t^k| + \delta}\right].$$

Using the definition of u^k and Morrey's estimate (4.31) we can bound the first addend by

$$\begin{aligned}\Big|u(s, X_t) - u^k(s, X_t^k)\Big| &\leq \fint_{t^n}^{t^{n+1}} \Big|u(t, X_t) - u(t, X_s^k)\Big| \,\mathrm{d}s \\ &\lesssim \fint_{t^n}^{t^{n+1}} \Big((M\nabla\overline{u})(t, X_t) + (M\nabla\overline{u})(t, X_s^k)\Big)|X_t - X_s^k| \,\mathrm{d}s\,.\end{aligned}$$

Plugging this estimate into I^n, we introduce the normalized Lebesgue measure

$$\mathrm{d}\omega_0(x) = \frac{\mathbb{1}_{\mathrm{D}}(x)}{|\mathrm{D}|} \,\mathrm{d}x$$

and using Hölder's inequality we can write

$$
\begin{aligned}
\mathrm{I}^n &\lesssim |\mathrm{D}| \int_{t^n}^{t^{n+1}} \fint_{t^n}^{t^{n+1}} \mathbb{E}_{\omega_0}\left[\Big((M\nabla\overline{u})(s,X_s) + (M\nabla\overline{u})(s,X^k_\tau)\Big)\frac{|X_s - X^k_\tau|}{|X_{t^n} - X^k_{t^n}| + \delta}|\theta^0|\right] \mathrm{d}\tau\ \mathrm{d}s \\
&\lesssim |\mathrm{D}| \int_{t^n}^{t^{n+1}} \fint_{t^n}^{t^{n+1}} \mathbb{E}_{\omega_0}\Big[|M\nabla\overline{u}(s,X_s)|^p + |M\nabla\overline{u}(s,X^k_\tau)|^p\Big]^{\frac{1}{p}} \\
&\quad \cdot \mathbb{E}_{\omega_0}\left[\left(\frac{|X_s - X^k_\tau|}{|X_s - X^k_s| + \delta}|\theta^0|\right)^q\right]^{\frac{1}{q}} \mathrm{d}\tau\ \mathrm{d}s \\
&\leq \int_{t^n}^{t^{n+1}} \fint_{t^n}^{t^{n+1}} \left(\int_\Omega |M\nabla\overline{u}(s,x)|^p\ \mathrm{d}\omega_s + \int_\Omega |M\nabla\overline{u}(s,x)|^p\ \mathrm{d}\omega^k_\tau\right)^{\frac{1}{p}} \\
&\quad \cdot \mathbb{E}_{\theta^0}\left[\left(\frac{|X_s - X^k_\tau|}{|X_s - X^k_s| + \delta}|\theta^0|\right)^q\right]^{\frac{1}{q}},
\end{aligned}
$$

where ω_t and ω^k_t are by the representation (A.9) solutions to (4.1) with initial datum ω_0 driven by u and u^k, respectively.

Now the L^p norm of the maximal function is directly controlled by the fundamental inequality for maximal functions (4.30). For the rest we can apply the elemental inequality,

$$
|X_s - X^k_\tau|^q \leq 2^{q-1}(|X_s - X^k_s|^q + |X^k_s - X^k_\tau|^q)
$$

and by the definition of the stochastic flow (4.36) we have for any $t, s \in [t^n, t^{n+1})$, $s \leq t$, the estimate

$$
\begin{aligned}
\mathbb{E}[|X_t - X_s|^q] &\lesssim \mathbb{E}\left[\left|\int_{t^n}^{t^{n+1}} u(s,X_s)\,\mathrm{d}s\right|^q\right] + (2\kappa)^{q/2}\mathbb{E}[|B_{t^{n+1}} - B_{t^n}|^q] \\
&\quad + \mathbb{E}\left[\left|\int_{t^n}^{t^{n+1}} n(X_s)\,\mathrm{d}L_s\right|^q\right] \\
&\leq \|u\|^q_\infty k^q + (2\kappa k)^{q/2} + k^q
\end{aligned}
$$

where we have used the Doob maximal martingale inequality for the Brownian motion (4.37) and the standard bound for the $\mathcal{F}_t$-adapted process L_t (A.4) together with the trivial property of the normal vector $\|n\|_{L^\infty} = 1$. Therefore, by means of these last two inequalities we can write

$$\begin{aligned}\mathbb{E}\left[\left(\frac{|X_s-X^k_\tau|}{|X_s-X^k_s|+\delta}\right)^q\right] &\le \mathbb{E}\left[\frac{2^{q-1}(|X_s-X^k_s|^q+|X^k_s-X^k_\tau|^q)}{(|X_s-X^k_s|+\delta)^q}\right]\\ &\lesssim \mathbb{E}\left[\frac{|X_s-X^k_s|^q}{|X_s-X^k_s|^q}+\frac{|X^k_s-X^k_\tau|^q}{\delta^q}\right]\\ &\lesssim 1+\frac{(\|u\|^q_\infty+1)k^q+(\kappa k)^{q/2}}{\delta^q}.\end{aligned}$$

Finally, noticing that $(1+x^q)^{1/q}\le 1+x$ and $(x^q+y^q)^{1/q}\le x+y$ for all $q>1$ and $x,y>0$, it yields the estimate for I^n,

$$\begin{aligned}\mathrm{I}^n &\lesssim \left(1+\frac{(\|u\|^q_\infty+1)k^q+(\kappa k)^{q/2}}{\delta^q}\right)^{1/q}\|\theta^0\|_{L^q}\int_{t^n}^{t^{n+1}}\|\nabla\overline{u}(s)\|_{L^p}\,\mathrm{d}s\\ &\lesssim \left(1+(\|u\|_\infty+1)\frac{k}{\delta}+\frac{\sqrt{\kappa k}}{\delta}\right)\|\theta^0\|_{L^q}\int_{t^n}^{t^{n+1}}\|\nabla\overline{u}(s)\|_{L^p}\,\mathrm{d}s.\end{aligned}\tag{4.39}$$

and hence by combining it with (4.38) and using that $u\in L^1(W^{1,p})$ we get the result stated by the Lemma. ■

At this point we collect the three discretization errors (time, initial data, vector-field) from Lemmas 4.4– 4.6. Since the Kantorovich–Rubinstein distance $\mathcal{D}_\delta(\cdot,\cdot)$ satisfies the triangle inequality we can just write now for any $t\in[t^m,t^{m+1})$ and any $m\in[\![0,N-1]\!]$,

$$\begin{aligned}\mathcal{D}_\delta(\theta(t),\theta_{k,h}(t)) \le \mathcal{D}_\delta(\theta(t),\theta(t^m))+\mathcal{D}_\delta(\theta(t^m),\theta^h(t^m))\\ +\mathcal{D}_\delta(\theta^h(t^m),\theta^{k,h}(t^m))+\mathcal{D}_\delta(\theta^{k,h}(t^m),\theta_{k,h}(t^m)),\end{aligned}$$

where $\theta^{k,h}$ is the unique solution to the advection-diffusion equation (4.1) with vector field u^k and initial datum θ^0_h. Notice that Lemma 4.4 and 4.5 yield control over the first two addends. In order to get the bound for the third addend we apply Lemma 4.6 with $\theta^0=\theta^0_h$, and hence we arrive to the expression

$$\mathcal{D}_\delta(\theta(t),\theta_{k,h}(t))\lesssim 1+\frac{h+k\|u\|_\infty+\sqrt{k\kappa}}{\delta}+\mathcal{D}_\delta(\theta^{k,h}(t^m),\theta_{k,h}(t^m)).\tag{4.40}$$

The last addend in (4.40) corresponds to the so-called truncation error or error caused by the scheme. We will concentrate on it in the next section.

4.3.2. Error due to the scheme

Since we already studied the errors coming from the discretization of the initial datum and vector field, we can consider now the continuous problem (4.1) with vector field u^k and initial datum θ_h^0. Also we can assume that $t = t^m$ for some $m \in [\![0, N]\!]$ such that we have $\theta(t, x) = \theta^{k,h}(t^m, x)$. However, for the sake of a clear notation, along this section we will write θ denoting $\theta^{k,h}$.

We want to study the distance $\mathcal{D}_\delta(\theta(t^m), \theta_{k,h}(t^m))$ and in order to do so it is more convenient to consider a piecewise linear temporal approximation of $\theta_{k,h}$ defined by

$$\hat{\theta}_{k,h}(t, x) = \frac{t - t^n}{k}\theta_K^{n+1} + \frac{t^{n+1} - t}{k}\theta_K^n \qquad \text{for a.e. } (t, x) \in [t^n, t^{n+1}) \times K$$

for all $K \in \mathcal{T}$ and all $n \in [\![0, N]\!]$. One can check that indeed for the time points of the mesh t^n with $n \in [\![0, N]\!]$ it holds $\theta_{k,h}(t^n) = \hat{\theta}_{k,h}(t^n)$ and hence no additional error term must be considered. This linear piecewise temporal approximation is particularly convenient because it is weakly differentiable and by construction it holds,

$$\partial_t \hat{\theta}_{k,h}(t, x) = \frac{\theta_K^{n+1} - \theta_K^n}{k} \quad \text{for a.e. } (t, x) \in [t^n, t^{n+1}) \times K.$$

Therefore we can directly apply (2.11) to obtain

$$\frac{d}{dt}\mathcal{D}_\delta(\theta, \hat{\theta}_{k,h}) = \int_{\mathrm{D}} \nabla\zeta \cdot u\theta \,\mathrm{d}x + \kappa \int_{\mathrm{D}} \zeta\Delta\theta \,\mathrm{d}x - \frac{1}{k}\sum_K \int_K \zeta(\theta_K^{n+1} - \theta_K^n) \,\mathrm{d}x.$$

where ζ represents the optimal Kantorovich potential associated to the distance $\mathcal{D}_\delta(\theta, \hat{\theta}_{k,h})$.

For the last term in the right hand side we can use the definition of the upwind scheme (4.21) in an analogous process to what it is done with the continuous part. Then, after integration over $[t^n, t^{n+1})$ we get

$$\mathcal{D}_\delta(\theta(t^{n+1}), \hat{\theta}_{k,h}(t^{n+1})) - \mathcal{D}_\delta(\theta(t^n), \hat{\theta}_{k,h}(t^n)) = \mathrm{I}^n + \mathrm{II}^n + \mathrm{III}^n + \mathrm{IV}^n \tag{4.41}$$

with

$$\mathrm{I}^n = \int_{t^n}^{t^{n+1}} \int_{\mathrm{D}} \nabla\zeta \cdot u(\theta - \theta_h^{n+1})\,\mathrm{d}x\ \mathrm{d}t, \tag{4.42}$$

$$\mathrm{II}^n = \int_{t^n}^{t^{n+1}} \int_{\mathrm{D}} \nabla\zeta \cdot u\theta_h^{n+1}\,\mathrm{d}x\ \mathrm{d}t + k\sum_K \zeta_K^n \sum_{L\sim K} |K|L|u_{KL}^n \frac{\theta_K^{n+1} + \theta_L^{n+1}}{2}, \tag{4.43}$$

$$\mathrm{III}^n = k\sum_K \zeta_K^n \sum_{L\sim K} |K|L||u_{KL}^n| \frac{\theta_K^{n+1} - \theta_L^{n+1}}{2}, \tag{4.44}$$

$$\mathrm{IV}^n = \kappa \int_{t^n}^{t^{n+1}} \sum_K \int_K \zeta \left(\Delta\theta - \sum_{L\sim K} \frac{|K|L|}{|K|} \frac{\theta_L^{n+1} - \theta_K^{n+1}}{d_{KL}} \right) \mathrm{d}x\ \mathrm{d}t, \tag{4.45}$$

where we use the notation

$$\zeta_K^n = \fint_{t^n}^{t^{n+1}} \fint_K \zeta dx\, dt.$$

We will study the contribution to the final error caused by the scheme analysing the four terms separately in the four following Lemmas.

Lemma 4.7 (Error from I^n)**.** *The first contribution to the error caused by the scheme is*

$$\sum_n \mathrm{I}^n \lesssim 1 + \frac{\sqrt{kT}\|u\|_\infty}{\delta}.$$

We will omit the proof of this Lemma for the sake of brevity because the argument is completely analogous to the one in [80, Lemma 7]: a combination of properties of the optimal transport distance, Morrey's inequality and stability estimates.

From now on our procedure here diverges from the techniques in [80], providing indeed the better convergence rate for the size of the mesh h.

Lemma 4.8 (Error from II^n)**.** *The second contribution to the error caused by the scheme is*

$$\sum_n \mathrm{II}^n \lesssim \frac{h}{\delta} \min\left\{ \|u\|_\infty \sqrt{\frac{T}{\kappa}}, \sqrt{\frac{T\|u\|_\infty}{h}} \right\}.$$

Proof. In order to proof the estimate for II^n first it is convenient to rewrite it in a more suitable way. Let us abuse the notation for the sake of a clear exposition of the results and write u^n for $u(t^n)$ and ζ^n for the average of ζ over the interval $[t^n, t^{n+1})$. With this notation for the first addend in II^n notice that

$$\int_K \nabla\zeta \cdot u^n\,\mathrm{d}x = \sum_{L\sim K} \int_{K|L} \zeta u^n \cdot n_{KL}\,\mathrm{d}\mathcal{H}^{d-1} - \int_K \zeta \nabla \cdot u^n\,\mathrm{d}x.$$

Meanwhile, since $u^n_{KL} = -u^n_{LK}$, for the second addend it holds

$$k\sum_K \zeta^n_K \sum_{L\sim K} |K|L|u^n_{KL}\frac{\theta^{n+1}_K+\theta^{n+1}_L}{2} = k\sum_K \theta^{n+1}_K \sum_{L\sim K} |K|L|u^n_{KL}\frac{\zeta^n_K-\zeta^n_L}{2}.$$

Therefore we can develop the whole term as follows

$$\begin{aligned}
\mathrm{II}^n &= k\sum_K \theta^{n+1}_K \sum_{L\sim K}\left[\int_{K|L}\zeta^n u^n\cdot n_{KL}\,\mathrm{d}\mathcal{H}^{d-1} - |K|L|u^n_{KL}\frac{\zeta^n_K+\zeta^n_L}{2}\right] \\
&\quad - k\sum_K \theta^{n+1}_K\int_K(\zeta^n-\zeta^n_K)\nabla\cdot u^n\,\mathrm{d}x \\
&= k\sum_K \theta^{n+1}_K \sum_{L\sim K}\int_{K|L}\zeta^n\left[(u-u^n_K) - \fint_{K|L}(u^n-u^n_K)\,\mathrm{d}\mathcal{H}^{d-1}\right]\cdot n_{KL}\,\mathrm{d}\mathcal{H}^{d-1} \\
&\quad + k\sum_K \theta^{n+1}_K \sum_{L\sim K}|K|L|u^n_{KL}\left(\fint_{K|L}\zeta^n\,\mathrm{d}\mathcal{H}^{d-1} - \frac{\zeta^n_K+\zeta^n_L}{2}\right) \\
&\quad - k\sum_K \theta^{n+1}_K\int_K(\zeta^n-\zeta^n_K)\nabla\cdot u^n\,\mathrm{d}x,
\end{aligned}$$

so that we call the three addends II^n_1, II^n_2 and II^n_3 respectively.

First, to estimate II^n_1, we can use the fact that ζ^n_K is constant to add and subtract on each $L\sim K$ a term of the form

$$\zeta^n_K\int_{K|L}(u^n-u^n_K)\cdot n_{KL}\,\mathrm{d}\mathcal{H}^{d-1}$$

such that we obtain

$$\begin{aligned}
\mathrm{II}^n_1 &= k\sum_K \theta^{n+1}_K \sum_{L\sim K}\int_{K|L}(\zeta^n-\zeta^n_K)(u-u^n_K)\cdot n_{KL}\,\mathrm{d}\mathcal{H}^{d-1} \\
&\quad + k\sum_K \theta^{n+1}_K \sum_{L\sim K}\int_{K|L}\zeta^n_K(u-u^n_K)\cdot n_{KL}\,\mathrm{d}\mathcal{H}^{d-1} \\
&\quad - k\sum_K \theta^{n+1}_K \sum_{L\sim K}\int_{K|L}\zeta^n\fint_{K|L}(u^n-u^n_K)\,\mathrm{d}\mathcal{H}^{d-1}\cdot n_{KL}\,\mathrm{d}\mathcal{H}^{d-1} \\
&\le 2k\sum_K \theta^{n+1}_K\|\zeta^n-\zeta^n_K\|_{L^\infty}\int_{\partial K}|u^n-u^n_K|\,\mathrm{d}\mathcal{H}^{d-1}.
\end{aligned}$$

Now on the one hand we use the Lipschitz condition of the Kantorovich potential, i.e.

for every $x \in K$ it holds

$$|\zeta(x) - \zeta_K| = \left| \fint_K (\zeta(x) - \zeta(y)) \, \mathrm{d}y \right| \le \|\nabla\zeta\|_{L^\infty} \fint_K |x - y| \, \mathrm{d}y \lesssim \frac{h}{\delta}. \tag{4.46}$$

On the other hand by means of the trace and the Poincaré inequality (4.9) we obtain

$$\int_{\partial K} |u^n - u_K^n| \, \mathrm{d}\mathcal{H}^{d-1} \lesssim \|\nabla u^n\|_{L^1(K)} + \frac{1}{h}\|u^n - u_K^n\|_{L^1(K)} \lesssim \|\nabla u^n\|_{L^1(K)}.$$

Therefore combining everything, summing over n and using Hölder's inequality we get the estimate,

$$\sum_n \mathrm{II}_1^n \lesssim \frac{kh}{\delta} \sum_n \sum_K \theta_K^{n+1} \|\nabla u^n\|_{L^1(K)} \le \frac{h}{\delta} \|\theta_{k,h}\|_{L^\infty(L^q)} \|\nabla u\|_{L^1(L^p)}.$$

For II_2^n instead we use again that $u_{KL}^n = -u_{LK}^n$ and hence we can rewrite the term as

$$\mathrm{II}_2^n = k \sum_K \frac{\theta_K^{n+1} - \theta_L^{n+1}}{2} \sum_{L \sim K} |K|L| u_{KL}^n \left(\fint_{K|L} \zeta^n \, \mathrm{d}\mathcal{H}^{d-1} - \frac{\zeta_K^n + \zeta_L^n}{2} \right),$$

so that by the Lipschitz property of ζ the last factor is bounded by

$$\fint_{K|L} \zeta^n \, \mathrm{d}\mathcal{H}^{d-1} - \frac{\zeta_K^n - \zeta_L^n}{2} \lesssim \fint_K \fint_{K|L} (\zeta^n(x) - \zeta(y)) \, \mathrm{d}\mathcal{H}^{d-1}(x) \, \mathrm{d}y \lesssim \frac{h}{\delta}.$$

Therefore, II_2^n is controlled then by a term of the form

$$\mathrm{II}_2^n \lesssim \frac{kh}{\delta} \|u\|_\infty \sum_K \sum_{K|L} |K|L| |\theta_K^{n+1} - \theta_L^{n+1}|,$$

and thus a direct application of (4.27) yields

$$\sum_n \mathrm{II}_2^n \lesssim \frac{h}{\delta} \min\left\{ \|u\|_\infty \sqrt{\frac{T}{\kappa}}, \sqrt{\frac{T\|u\|_\infty}{h}} \right\}.$$

Finally for the third addend II_3^n we make use again of the Lipschitz property of ζ from (4.46) and we bound the divergence of the vector field u^n by its gradient and some dimension dependant constant such that we obtain

$$\mathrm{II}_3^n \lesssim \frac{kh}{\delta} \sum_K |\theta_K^{n+1}| \int_K |\nabla u^n| \, \mathrm{d}x \le \frac{kh}{\delta} \|\theta_h^{n+1}\|_{L^q} \|\nabla u^n\|_{L^p}.$$

After summation in n we get a bound analogous to the bound that we got for II_1^n and thus all three addends in II^n are controlled by the factor stated in the claim of the Lemma. ■

Lemma 4.9 (Error from III^n)**.** *The third contribution to the error caused by the scheme is*

$$\sum_n \mathrm{III}^n \lesssim \frac{h}{\delta}\min\left\{\|u\|_\infty\sqrt{\frac{T}{\kappa}}, \sqrt{\frac{T\|u\|_\infty}{h}}\right\}.$$

Proof. The proof of this Lemma follows a similar strategy to what has been performed in the previous one. First of all notice that this time $|u_{KL}^n| = |-u_{KL}^n|$ and hence we can symmetrize III^n as

$$\mathrm{III}^n = k\sum_K\sum_{L\sim K}|K|L||u_{KL}^n|\frac{\zeta_K^n - \zeta_L^n}{2}\frac{\theta_K^{n+1} - \theta_L^{n+1}}{2}.$$

Since the Kantorovich potential is Lipschitz we have the bound

$$|\zeta_K^n - \zeta_L^n| \leq \fint_K\fint_L|\zeta^n(x) - \zeta^n(y)|\,\mathrm{d}x\,\mathrm{d}y \leq \|\nabla\zeta\|_{L^\infty}\fint_K\fint_L|x-y|\,\mathrm{d}x\,\mathrm{d}y \lesssim \frac{h}{\delta}$$

and hence

$$\mathrm{III}^n \lesssim \frac{kh}{\delta}\|u\|_\infty\sum_K\sum_{L\sim K}|K|L||\theta_K^n - \theta_L^n|.$$

After summation in n, by means of the BV estimate (4.27) as before we obtain the statement of the Lemma. ■

Finally, to study the contribution made by the diffusion term we will follow a similar technique to what it is done in Lemma 4.4 but adapting it now to the setting of a finite volume scheme. In order to make this suitable approximation of the Laplacian we need to argue as follows.

Given an admissible tessellation $\mathcal{T}$ of D and two neighbouring cells $K, L \in \mathcal{T}$ we define a diffeomorphism $\phi_{KL} : K \to L$ with constant Jacobian derivative, what means

$$J\phi_{KL} \equiv |\det\nabla\phi_{KL}| = \frac{|L|}{|K|}$$

such that the mass is preserved. Since all the admissible cells are convex the existence of this map is guaranteed, for instance consider an appropriate Brenier map [17, 64] or some other analogous construction [2]. Then, using this diffeomorphism we can define a *finite-volume-based* approximation of the Laplacian such as

$$\Delta^h f(x) = \sum_{L\sim K} \frac{|K|L|}{|K|} \frac{f\circ\phi_{KL}(x) - f(x)}{d_{KL}} \quad \text{for a.e. } x\in K \text{ and all } K\in\mathcal{T}.$$

Indeed, for sufficiently regular functions f it holds $\lim_{h\to 0}\Delta^h f = \Delta f$. This will be a key instrument in the proof of the next and last Lemma.

Lemma 4.10 (Error from IV^n). *The fourth term does not contribute to the error caused by the scheme, that is*

$$\sum_n \mathrm{IV}^n \le 0.$$

Proof. To prove this result we follow an adapted version of the technique used in Lemma 4.4 that the authors explain in more detail in [72]. This technique in turn comes inspired by [46]. Let us start by considering an approximation of the Laplacian as explained in the previous paragraphs. By means of Δ^h we can also define an approximation to IV^n as follows,

$$\begin{aligned}\frac{1}{\kappa}\mathrm{IV}^n_h &= \int_{t^n}^{t^{n+1}}\sum_K\int_K \zeta\left(\Delta^h\theta - \sum_{L\sim K}\frac{|K|L|}{|K|}\frac{\theta_L^{n+1}-\theta_K^{n+1}}{d_{KL}}\right)\mathrm{d}x\,\mathrm{d}t\\ &= \int_{t^n}^{t^{n+1}}\sum_K\int_K \zeta(x)\left(\sum_{L\sim K}\frac{|K|L|}{|K|}\frac{\theta\circ\phi_{KL}(x)-\theta(x)}{d_{KL}} - \sum_{L\sim K}\frac{|K|L|}{|K|}\frac{\theta_L^{n+1}-\theta_K^{n+1}}{d_{KL}}\right)\mathrm{d}x\,\mathrm{d}t.\end{aligned}$$

Notice that since $\zeta\in W^{1,\infty}$ and $\theta\in W^{1,1}$ it holds

$$\lim_{h\to 0}\mathrm{IV}^n_h = \mathrm{IV}^n,$$

and thus it is enough to study the approximation IV^n_h instead of IV^n.

Through a convenient change of variables $y=\phi_{KL}(x)$ on the first addend that we can make because ϕ_{KL} is a diffeomorphism, it yields

$$\begin{aligned}\frac{1}{\kappa}\mathrm{IV}^n_h &= \int_{t^n}^{t^{n+1}}\sum_K\sum_{L\sim K}\frac{|K|L|}{|K|}\frac{1}{d_{KL}}\int_K\zeta(x)\left[(\theta\circ\phi_{KL}(x)-\theta_L^{n+1}) - (\theta(x)-\theta_K^{n+1})\right]\mathrm{d}x\,\mathrm{d}t\\ &= \int_{t^n}^{t^{n+1}}\sum_K\sum_{L\sim K}\frac{|K|L|}{|L|}\frac{1}{d_{KL}}\int_L\zeta\circ\phi_{KL}^{-1}(y)(\theta(y)-\theta_L^{n+1})\,\mathrm{d}y\,\mathrm{d}t\\ &\quad - \int_{t^n}^{t^{n+1}}\sum_K\sum_{L\sim K}\frac{|K|L|}{|K|}\frac{1}{d_{KL}}\int_K\zeta(x)(\theta(x)-\theta_K^{n+1})\,\mathrm{d}x\,\mathrm{d}t,\end{aligned}$$

where we have used that the Jacobian derivative of ϕ_{KL} is constant and equals $|L|/|K|$.

Then notice that by definition ζ is the optimal Kantorovich potential for the distance between θ and $\theta_{k,h}$, i.e.

$$\sum_K \int_K \zeta(x)(\theta(x) - \theta_K^{n+1})\,\mathrm{d}x = \mathcal{D}_\delta(\theta, \theta_{k,h}).$$

Furthermore, the optimal ζ is taken as the supremum over a set of functions where $\zeta \circ \phi_{KL}^{-1}$ also belongs to. Therefore it holds

$$\sum_K \int_K \zeta \circ \phi_{KL}^{-1}(x)(\theta(x) - \theta_K^{n+1})\,\mathrm{d}x \le \mathcal{D}_\delta(\theta, \theta_{k,h})$$

and hence, after relabelling in a suitable way

$$\frac{1}{\kappa}\,\mathrm{IV}_h^n \le \int_{t^n}^{t^{n+1}} \sum_K \sum_{L\sim K} \frac{|K|L|}{|K|}\frac{1}{d_{KL}}(\mathcal{D}_\delta(\theta, \theta_{k,h})|_K - \mathcal{D}_\delta(\theta, \theta_{k,h})|_K)\,\mathrm{d}t = 0, \tag{4.47}$$

where we denote by $\mathcal{D}_\delta(\theta, \theta_{k,h})|_K$ the restriction of the distance $\mathcal{D}_\delta(\theta, \theta_{k,h})$ to the subset $K \subset \mathrm{D}$. Since (4.47) holds uniformly in h, it yields that in the limit IV^n does not contribute to the error caused the scheme. ■

Proof of Theorem 4.1. Finally we can get the result on Theorem 4.1 with a straightforward combination of Lemmas 4.4–4.10. The first three of them already yield the intermediate estimate (4.40). For the remaining term we just notice that by definition $\theta^{k,h}(0) = \theta_{k,h}(0) = \theta_h^0$ and thus we can sum on (4.41) so that

$$\begin{aligned}\mathcal{D}_\delta(\theta^{k,h}(t^m), \theta_{k,h}(t^m)) &= \sum_{n=0}^{m}(\mathrm{I}^n + \mathrm{II}^n + \mathrm{III}^n + \mathrm{IV}^n)\\ &\lesssim 1 + \frac{h}{\delta}\min\left\{\|u\|_\infty\sqrt{\frac{T}{\kappa}}, \sqrt{\frac{T\|u\|_\infty}{h}}\right\} + \frac{\sqrt{kT}\|u\|_\infty}{\delta}.\end{aligned}$$

Combining this with (4.40) we get the estimate on Theorem 4.1. ■

5. Ergodicity and mixing results with random vector fields

Chapter summary

In this work we consider a passive scalar advected by a velocity field that evolves via a parameter that follows a random path. We prove that, provided that certain operator involving the vector field is hypocoercive, solution to the transport problem driven by such stochastic vector field is ergodic. As a consequence we obtain that the vector field mixes any initial configuration for the passive scalar averaged on the noise (annealed mixing). In addition, we are able to provide some examples of vector fields that produce annealed mixing in different types of bounded domains, including a randomly driven smooth vortex on a disc.

5.1. Introduction

In this chapter we are concerned with the transport of a passive scalar $\theta : \mathrm{D} \to \mathbb{R}$ under the action of a divergence-free vector field $u : (0, \infty) \times \mathrm{D} \to \mathbb{R}^d$. The problem is posed in a domain $\mathrm{D} \subset \mathbb{R}^d$ that throughout the whole chapter we will consider to be convex and bounded, and we will distinguish between the cases when D has boundary and when it has not. In case $\partial\mathrm{D} \neq \varnothing$, the boundary will be sufficiently smooth.

On the one hand we consider the transport equation that defines the time evolution of the passive scalar,

$$\begin{cases} \partial_t\theta + u \cdot \nabla\theta &= 0 \quad \text{in } (0,\infty) \times \mathrm{D}, \\ \theta(0,\cdot) &= \theta_0 \quad \text{in } \mathrm{D}. \end{cases} \tag{5.1}$$

where notice that the drift term verifies $u \cdot \nabla\theta = \nabla \cdot (u\theta)$ due to the divergence free assumption on u.

On the other hand we consider the advection-diffusion equation, where the evolution of the passive scalar is determined by the action of the vector field and the effect of diffusion. We denote the passive scalar for this problem by θ^κ. We assume $\kappa > 0$ to be the diffusion coefficient, and therefore the advection-diffusion equation takes the following form,

$$\begin{cases} \partial_t \theta^\kappa + u \cdot \nabla \theta^\kappa &= \kappa \Delta \theta^\kappa \quad \text{in } (0,\infty) \times \mathrm{D}, \\ \theta^\kappa(0,\cdot) &= \theta_0 \quad \text{in } \mathrm{D}. \end{cases} \tag{5.2}$$

If $\mathrm{D} \subset \mathbb{R}^d$ has a boundary, we consider in addition homogeneous Neumann boundary conditions,

$$n \cdot \nabla \theta^\kappa = 0 \quad \text{on } (0,\infty) \times \partial \mathrm{D}.$$

This is the only condition for the boundary of D to be consider since we will always assume that the vector field is always tangent to it, namely $u \cdot n = 0$ on ∂D.

The main goal of the chapter is the derivation of ergodicity and mixing results for passive scalars θ and θ^κ when they are driven by stochastic vector fields that satisfy certain properties. Moreover, even on a random setting, we want to address these problems from a purely analytical perspective.

Mixing of passive scalars has been a topic of great interests for the community in recent years. Due to the very many applications, physicists have been interested in the mixing problem since decades ago. However, from a more rigorous or mathematical perspective, the mixing problem has gained more interest only recently due to a result on the cost of rearrangements by Bressan, [19]. In this work, the author introduces the concept of *geometric mixing scale* and states a conjecture that has only been partially solved yet in [29]. Bressan's work opened the door to the geometrical study of the mixing properties of vector fields, and shortly after a functional analytical perspective of the mixing problem was introduced in [38, 63]. In those works, the authors consider the decay of a negative Sobolev norm as a measure of the degree of mixedness, the so-called *functional mixing scale*, see Section 2.4 for more information about the mixing problem. The latter is precisely the notion that we want to consider here.

Let θ be a solution to the transport equation (5.1) or (5.2). We say that the vector field u mixes the passive scalar if

$$\|\theta(t)\|_{\dot{H}^{-s}(\mathrm{D})} \to 0 \quad \text{as } t \to \infty$$

for some $s > 0$. Along the chapter we will use the particular case $s = 1$ as a measure of mixing since the $\dot{H}^{-1}-$norm scales like [Length] (in general the $\dot{H}^{-s}-$norm scales like [Length]^s). This observation comes from the fact that we define the negative Sobolev norms via

$$\|f\|^2_{\dot{H}^{-s}(\mathrm{D})} = \sum_{k>0} \lambda^{-s} \langle \phi_k, f \rangle^2,$$

where $\{\phi_k\}_{k>0}$ is an orthonormal basis for mean free functions in $L^2(\mathrm{D})$ given by the Neumann laplacian, and λ_k are the corresponding eigenvalues. See Section 2.4 for further details.

In the literature there are already a number of examples of vector fields that mix any initial configuration in $L^\infty \cap H^s$ with $s > 0$. If one considers deterministic vector fields that are smooth, or at least sufficiently regular to have well-posedness of the transport equation (5.1), then it is known that these cannot mix faster than exponential [29, 54, 82]. However there are some well-known examples of vector fields for which solutions to (5.1) satisfy

$$\|\theta(t)\|_{\dot{H}^{-1}(\mathrm{D})} \lesssim \|\theta_0\|_{H^1(\mathrm{D})} e^{-\gamma t},$$

for some $\gamma > 0$ and all $t > 0$, see e.g. [1, 40, 96]. There are in addition important examples of vector fields that are known to mix polynomially in time but not faster, e.g. [26, 30]. This is particularly relevant for our purposes because vector fields like steady point vortices or shear flows in a deterministic setting fall into the class of polynomial mixers.

For random vector fields, the business of (almost-sure) mixing has been very recently addressed by the ground-breaking work of Bedrossian, Blumenthal and Punshon-Smith, [6, 7, 8, 9]. They designed a program about Lagrangian chaos, exponential mixing and enhanced dissipation with vector fields u that are solution to a stochastically forced version of Navier-Stokes. Their methods already had applications for alternating shear flows with random phase or random durations [12, 27]. All these results are based on a dynamical perspective of the transport equation, where the estimates are derived in terms of stochastic particle trajectories and using methods from the theory of Markov processes and random dynamical systems.

In this work we introduce a class of vector fields that are defined on

$$u : \mathrm{D} \times \mathrm{D}' \to \mathbb{R}^d,$$

where the new domain $\mathrm{D}' \subset \mathbb{R}^d$ is bounded and may or may not have a boundary. We denote by y the new variable in D', and we evaluate the vector field $u = u(x, y)$ on the second entry with a stochastic process $\{Y_t\}_{t\geq 0}$, that will be the source of randomness for our problem (5.1). Therefore the vector field $u = u(x, Y_t)$ depends on $(t, x) \in (0, \infty) \times \mathrm{D}$ and, implicitely also on the noise $\omega \in \Omega$.

The specific process that we want to consider is described as follows. Fix a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$, then we define the stochastic process $\{Y_t\}_{t\geq 0}$ to be a *reflecting Brownian motion* in D', namely the solution to the SDE

$$\begin{cases} \mathrm{d}Y_t &= \sqrt{2\nu}\,\mathrm{d}B_t - n(Y_t)\,\mathrm{d}L_t, \quad t>0, \\ Y_0 &= \mathrm{id}, \end{cases} \tag{5.3}$$

where $\nu > 0$, $\{B_t\}_{t\geq 0}$ is a standard Brownian motion in $\mathbb{R}^d$, $\{L_t\}_{t\geq 0}$ is a local time of the process Y_t that is only activated when the process hits the boundary $Y_t \in \partial\mathrm{D}'$, and $n(y)$ is the outer normal vector the boundary at position $y \in \partial\mathrm{D}'$. As usual, if the domain D' does not have a boundary, then $Y_t = B_t$. We elaborate more on reflecting Brownian motion in subsequent sections and Appendix A.1.

We want to describe the evolution of the passive scalar θ under the action of a vector field with this specific form. For now let us assume that we only have transport, $\kappa = 0$, so θ solves equation (5.1). We can use the Lagrangian perspective to formulate the transport equation (5.1) incorporating the definition of u and its dependence on the Brownian motion. Namely we obtain,

$$\begin{cases} \mathrm{d}X_t &= u(X_t, Y_t)\,\mathrm{d}t & \text{in } (0,\infty)\times\mathrm{D}, \\ \mathrm{d}Y_t &= \sqrt{2\nu}\,\mathrm{d}B_t - n(Y_t)dL_t & \text{in } (0,\infty)\times\mathrm{D}'. \end{cases} \tag{5.4}$$

With this equation in mind, we transform it into a PDE that describes the evolution of the passive scalar

$$f(t,\cdot,\cdot) = \mathbb{E}(X_t, Y_t)_{\#} f_0,$$

by means of Kolmogorov equation or Feynman-Kac formula, see Appendix A.1. Taking into account that the Brownian motion produces a laplacian operator in the PDE framework, we obtain that the scalar function $f : (0,\infty)\times\mathrm{D}\times\mathrm{D}' \to \mathbb{R}$ solves the equation,

$$\begin{cases} \partial_t f + u\cdot\nabla_x f &= \nu\Delta_y f & \text{in } (0,\infty)\times\mathrm{D}\times\mathrm{D}', \\ n_y\cdot\nabla_y f &= 0 & \text{on } (0,\infty)\times\mathrm{D}\times\partial\mathrm{D}', \\ f(0,\cdot,\cdot) &= f_0 & \text{in } \mathrm{D}\times\mathrm{D}'. \end{cases} \tag{5.5}$$

Since we are interested in the problem for the marginal θ defined in D, we assume an initial configuration for f of the form $f_0 = \theta_0 \otimes \rho$, where θ_0 is the initial datum for the original problem (5.1), and ρ corresponds to the initial law that we impose on the Brownian motion $\{B_t\}_{t\geq 0}$, i.e. $\rho = \mathrm{law}\, Y_0$.

Notice that in the original problem (5.1), the passive scalar θ depends on the randomness $\omega \in \Omega$. However, the rewriting that we performed by using the properties of u yields a purely deterministic PDE (5.5) and hence we can apply all the very well-known machinery about long time behaviour of solutions.

We will assume along the whole chapter that the vector field u and the initial law for stochastic process $\{Y_t\}_{t\geq 0}$ have the following regularity:

$$u \in C^2(\mathrm{D} \times \mathrm{D}'), \quad \rho = \operatorname{law} Y_0 \in (\mathcal{P} \cap W^{1,\infty})(\mathrm{D}'). \tag{5.6}$$

Then, we derive results concerning the ergodicity of solution to (5.1) with this random vector field following a purely analytical perspective, namely we only need to study the decay of the solutions to equation (5.5). More in detail, the result that we prove is that if the operator defined in (5.5) is *hypocoercive*, i.e. if the H^1 norm of f decays exponentially fast, then solutions to the marginal problem (5.1) are exponentially ergodic: $\|\mathbb{E}\theta(t)\|_{L^2} \to 0$ with an exponential rate.

Some important assumption that we need to make on the initial datum θ_0 is that it is mean free. This is not a very restrictive condition since given any θ_0, one can always define $\tilde{\theta}_0 = \theta_0 - \fint \theta_0 \, \mathrm{d}x$, and the solution to the transport equation remains invariant. This is a by-product of the fact that the transport equation conserves mass, as explained in Section 2.1. In addition, we need assume the some integrability for the initial datum, so that the hypotheses assumed for θ_0 are

$$\theta_0 \in (L^\infty \cap H^1)(\mathrm{D}), \quad \fint_{\mathrm{D}} \theta_0 \, \mathrm{d}x = 0. \tag{5.7}$$

These conditions are not very surprising since what are concerned with a mixing problem, and thus in order to measure the decay of negative Sobolev norms we need to assume certain Sobolev regularity for the initial datum.

In this regard, the first result that we present here reads as follows.

Theorem 5.1 (Exponential ergodicity)**.** *Let the vector field $u \in C^2(\mathrm{D} \times \mathrm{D}')$ satisfy the assumptions* (5.6)*. Let $\theta_0 \in H^1(\mathrm{D})$ and $\nu > 0$. If the deterministic solution to* (5.5) *satisfies*

$$\|f(t)\|_{H^1(\mathrm{D}\times\mathrm{D}')} \lesssim \|f(0)\|_{H^1(\mathrm{D}\times\mathrm{D}')} e^{-\beta t},$$

for all $t \geq 0$, namely the operator in (5.5) *is hypocoercive, then there holds*

$$\|\mathbb{E}\theta(t)\|_{L^2(\mathrm{D})} \lesssim \|\theta_0\|_{H^1(\mathrm{D})} e^{-\beta t}$$

for all $t > 0$.

A straightforward consequence of Theorem 5.1is that the vector field is mixing exponentially fast any initial configuration in $H^1 \cap L^\infty$, after averaging on the noise (annealed mixing),

$$\|\mathbb{E}\theta(t)\|_{\dot{H}^{-1}(\mathrm{D})} \lesssim \|\theta_0\|_{H^1(\mathrm{D})} e^{-\beta t}$$

for all $t > 0$. This result might be surprising at first glance since we know equation (5.1) conserves the L^2 norm, however the noise averaging procedure plays a crucial role to obtain the decay.

In addition, we also provide some examples of ergodic solutions with random vector fields of the form $u = u(\cdot, Y_t)$ in different types of bounded domains. In particular we consider (non-alternating) shear flows with a random phase in the torus $\mathbb{T}^2$, and a vortex on the unit disc $B_1(0)$ that is randomly moving on a subset of the disc $B_1(0)$. Both examples are relevant due to the fact that, as it has been pointed out before, without the random phase or the random movement both the shear flows and the point vortex can only mix polynomially fast.

What it is proved in Theorem 5.1 is not yet a result concerning almost-sure exponential mixing, only ergodicity and thus annealed mixing which is a standard first step needed for the almost-sure mixing, see [7, 12]. At the present moments the authors conjecture that a similar result can be obtained for the pointwise mixing studying the hypocoercivity of analogous operator, but that is a result out of the scope of this thesis.

This chapter is organized as follows: In Section 5.2 an overview of the particle trajectories perspective for the mixing problem is given. Some basic theory about random dynamical system and the relevant Markov processes that must be studied for the mixing problem is introduced. In Section 5.3 the corresponding Eulerian perspective is introduced, that allows for PDE analysis to be used in order to come up with the mixing estimates. In addition, in Section 5.3.1 the notion of hypocoercivity is presented, since it will be relevant for providing examples. Section 5.4 deals with the proof of Theorem 5.1, and Sections 5.5 and 5.6 showcase examples of vector fields that produce exponential ergodicity in different types of bounded domains.

5.2. Lagrangian perspective

In this section we will give an overview about the Lagrangian perspective of the transport problem (5.1) and how ergodicity and mixing results for this equations can be studied from the particle trajectories viewpoint. This is the preferred perspective in the literature, as of today, in order to study ergodicity and almost-sure mixing properties of stochastic vector fields since one can make use of all the well-known theory of random dynamical systems and ergodicity of Markov processes.

Recall that the so-called Lagrangian perspective of the transport equation deals with the trajectories of the particles $X_t : \mathrm{D} \to \mathrm{D}$, which are solutions to the differential equation

$$\mathrm{d}X_t = u(t, X_t)\,\mathrm{d}t, \quad X_0 = \mathrm{id}\,.$$

This is an ODE if u is deterministic, and a SDE if u is stochastic. In this chapter we are concerned precisely with the stochastic case.

The study of mixing properties of stochastic vector fields from a Lagrangian perspective have gained a huge importance in the mathematical community in recent years. In the seminal work by Bedrossian, Blumenthal and Punshon-Smith, see [6, 7, 8], the authors consider a vector field that is solution to a stochastically forced Navier-Stokes equation. After this first ground-breaking contribution there have been other works that follow a similar strategy to prove ergodicity and mixing with different examples of random vector fields. Some of the most remarkable have been the contributions by Blumenthal, Coti Zelati, Gvalani and Cooperman, see [12, 27], where the authors study alternating shear flows with random phases and random durations.

The general setting concerns a bounded domain $\mathrm{D} \subset \mathbb{R}^d$ and a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, we define a vector field u that depends on time, space and the noise $\omega \in \Omega$ as in,

$$\begin{aligned} u: \quad (0,\infty) \times \mathrm{D} \times \Omega \quad &\rightarrow \quad \mathbb{R}^d \\ (t,x,\omega) \quad &\mapsto \quad u(t,x,\omega). \end{aligned}$$

Since we are interested in divergence free transport, we will always assume that

$$\nabla_x \cdot u(t,x,w) = 0 \quad \text{for all } (t,x,\omega) \in (0,\infty) \times \mathrm{D} \times \Omega.$$

With this stochastic vector field one can construct a stochastic flow, a solution to the transport SDE that will also depend on the noise $\omega \in \Omega$ via the Itô integral,

$$X_t^\omega(x) = x + \int_0^t u(s, X_s^\omega(x), \omega)\, \mathrm{d}s.$$

Thus, the central object of our study will be the stochatic process $\{X_t\}_{t\geq 0}$. A first key idea is to show that this process defines a *random dynamical system* in $(0,\infty) \times \mathrm{D} \times \Omega$. For this to hold, some special properties of u are needed. We will elaborate more on random dynamical systems in Section 5.2.1.

A second key idea is to show that we can define a Markov process $\{P_t\}_{t\geq 0}$ as follows,

$$P_t(x, A) = \mathbb{P}[X_t \in A \mid X_0 = x], \quad \text{for } t > 0, x \in \mathrm{D}, A \in \mathcal{F}. \tag{5.8}$$

Namely the object $P_t(x, A)$ gives information about the probability for the particle X_t to be in the Borel set $A \subseteq \mathrm{D}$ at time $t > 0$, provided that initially it was in $x \in \mathrm{D}$. This semigroup is usually referred in the literature as the *one-point process*.

Markov processes have some properties that are very helpful when dealing with the dynamics of the flow, for some basic definitions and key results about Markov processes and ergodicity of their invariant measures see Section 2.3.

The main argument in this line consists on proving that P_t has a unique invariant measure in D, the normalized Lebesgue measure, which yields that the Lebesgue measure is ergodic. If we want to prove exponential or geometric ergodicity we need to use some extra properties of u, via e.g. Harris' Theorem, see [52, 67]. In such case we obtain that for all $x \in \mathrm{D}$, the invariant measure is exponentially attracting the measure P_t, namely there exists $\alpha > 0$ such that

$$\left| P_t\phi(x) - \fint_{\mathrm{D}} \phi(z)\,\mathrm{d}z \right| \lesssim e^{-\alpha t}$$

for all $\phi \in L^\infty(\mathrm{D})$ and all $t > 0$.

Proving geometric ergodicity for the Markov process (5.8) is already a first interesting result since it implies *annealed mixing*, i.e. mixing averaged over all the possible noise realizations. However if the purpose is to study almost-sure exponential mixing, this is only the first step. In particular one needs to define two more Markov processes with the flow $\{X_t\}_{t\geq 0}$. On the one hand we define the *projective process*

$$\hat{P}_t((x,v),A) = \mathbb{P}\left[\left(X_t(x), \frac{(\nabla_x X_t)v}{|(\nabla_x X_t)v|}\right) \in A\right], \tag{5.9}$$

where $t > 0$, $(x,v) \in \mathrm{D} \times \mathbb{S}^1$ and $A \in \mathcal{F} \otimes \mathcal{B}(\mathbb{S}^1)$. On the other hand we define the *two-point process*

$$P_t^{(2)}((x_1,x_2),A) = \mathbb{P}\left[(X_t(x_1), X_t(x_2)) \in A\right], \tag{5.10}$$

where $t > 0$, $(x_1,x_2) \in \mathrm{D} \times \mathrm{D} \setminus \{(x,x) : x \in \mathrm{D}\}$ and $A \in \mathcal{F} \otimes \mathcal{F}$. This is precisely the key object to prove exponential mixing, since it is concerned with the the evolution of two particles that start at different positions in the domain D, and indeed it is proved in [7] that if $P_t^{(2)}$ has an invariant measure that is geometrically ergodic, then u mixes exponentially fast every initial datum $\theta_0 \in (H^1 \cap L^\infty)(\mathrm{D})$ almost surely.

Notice that $P_t^{(2)}$ is defined in a space that is not compact, $\mathrm{D} \times \mathrm{D} \setminus \{(x,x) : x \in \mathrm{D}\}$, therefore proving geometric ergodicity is more involved. The main strategy consists on:

1. Proving that P_t is geometrically ergodic.

2. Proving that the *top Lyapunov exponent* is strictly positive, namely

$$\lambda_+(x,\omega) = \lim_{t\to\infty} \frac{1}{t} \log |\nabla_x X_t^\omega| > 0$$

   for $\mathbb{P}-$a.e. $\omega \in \Omega$ and a.e. $x \in \mathrm{D}$.

3. Proving that $\hat{P}_t$ is geometrically ergodic.

With these results in hand, the fourth and final step is to prove that $P_t^{(2)}$ is geometrically ergodic, modulo some weight function V that satisfies certain drift conditions. For further details regarding this method see [7, 12].

For our specific problem we will consider a vector field $u = u(t, x, \omega)$ whose time and noise dependence are both encoded in a random parameter Y_t^ω that solves the SDE associate to a Brownian motion. If the domain D considered has a boundary, then the appropriate stochastic process to consider is the *reflecting Brownian motion*,

$$\begin{cases} \mathrm{d}Y_t &= \mathrm{d}B_t - n(Y_t)\,\mathrm{d}L_t, \quad t > 0, \\ Y_0 &= \mathrm{id}, \end{cases}$$

where $\{B_t\}_{t\geq 0}$ is a standard Brownian motion in $\mathbb{R}^d$, $\{L_t\}_{t\geq 0}$ is non-decreasing with $L_0 = 0$, and it holds

$$\int_0^t |n(Y_s)|\,\mathrm{d}L_s < \infty, \quad \text{and} \quad \int_0^t \chi_{\{Y_s \notin \partial \mathrm{D}'\}}\,\mathrm{d}L_s = 0$$

where $n(y)$ is the outer unit vector to the boundary of D' at position $y \in \partial\mathrm{D}'$. For more details about the reflecting Brownian motion see Section A.1 in the Appendix.

With this specific choice of randomness we can indeed show that the flow $\{X_t\}_{t\geq 0}$ forms a random dynamical system over some suitable metric dynamical system. We elaborate on that in the next Section.

5.2.1. Random dynamical systems

In this section we will give details on how to understand the flow $\{X_t\}_{t\geq 0}$ produced by the vector field u as a random dynamical system. As we exposed before, fixing a canonical filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we want to consider a reflecting Brownian motion in D',

$$\mathrm{d}Y_t = \sqrt{2\nu}\,\mathrm{d}B_t - n(Y_t)\,\mathrm{d}L_t, \quad Y_0 = \mathrm{id}\,. \tag{5.11}$$

Then we define a random vector field $u(x, Y_t^\omega)$ on a domain D that depends on the position $x \in \mathrm{D}$, and such that it has all the time and noise dependence encoded on the process Y_t^ω. With this stochastic vector field we consider the transport equation (5.1), or in Lagrangian form,

$$\mathrm{d}X_t = u(X_t, Y_t)\,\mathrm{d}t, \quad X_0 = \mathrm{id}\,. \tag{5.12}$$

This is a random flow whose randomness source is encoded on the Y_t-dependence of u. Therefore we can look at the flow $\{X_t\}_{t\geq 0}$ as a mapping

$$\begin{array}{rccc} X: & (0,\infty)\times\Omega\times \mathrm{D}\times\mathrm{D}' & \to & \mathrm{D} \\ & (t,\omega,x,y) & \mapsto & X_t^{\omega,y}(x), \end{array}$$

defined, via Itô integral, by

$$X_t^{\omega,y}(x) = x + \int_0^t u(X_s^{\omega,y}(x), Y_s^{\omega}(y))\,\mathrm{d}s.$$

Mostly in this chapter we will omit the superscripts ω, y in order to have a better notation, but for this section is more convenient keeping track of them explicitly.

The main purpose of the section is to prove that this mapping defines indeed a random dynamical system over a suitable metric dynamical system. First of all let us define what is a random dynamical system, for further details see the monograph [5].

Definition 5.1. Consider a probability space $(\Omega,\mathcal{F},\mathbb{P})$ and a mapping $\sigma:(0,\infty)\times\Omega\to\Omega$ such that

1. $\sigma_t:\Omega\to\Omega$ is measure preserving, i.e. $\mathbb{P}=(\sigma_t)_{\#}\mathbb{P}$, for every $t\geq 0$;
2. $\sigma_0=\mathrm{id}_\Omega$; and
3. $\{\sigma_t\}_{t\geq 0}$ has the semigroup property, i.e. $\sigma_{t+s}=\sigma_t\circ\sigma_s$ for all $t,s\geq 0$.

Then $(\Omega,\mathcal{F},\mathbb{P},\{\sigma_t\}_{t\geq 0})$ is called a *metric dynamical system.*

Definition 5.2. A *continuous random dynamical system* on a measurable space $(\mathrm{D},\mathcal{B}(\mathrm{D}))$ over the metric dynamical system $(\Omega,\mathcal{F},\mathbb{P},\{\sigma_t\}_{t\geq 0})$ is a mapping

$$\begin{array}{rccc} \varphi: & (0,\infty)\times\Omega\times \mathrm{D} & \to & \mathrm{D} \\ & (t,\omega,x) & \mapsto & \varphi_t^{\omega}(x), \end{array}$$

with the following properties,

1. φ is $(\mathcal{B}((0,\infty))\otimes\mathcal{B}(\mathrm{D})\otimes\mathcal{F},\mathcal{B}(\mathrm{D}))-$measurable;
2. *the cocycle property*: the mapping $\varphi_t^\omega:\mathrm{D}\to\mathrm{D}$ forms a cocycle over $\{\sigma_t\}_{t\geq 0}$, i.e.

   $$\varphi_0^\omega=\mathrm{id}_\mathrm{D},\quad \varphi_{t+s}^\omega(x)=\varphi_s^{\sigma_t(\omega)}(\varphi_t^\omega(x)),$$

   for every $s,t\geq 0$ and $\mathbb{P}-$almost all $\omega\in\Omega$;

3. for each $\omega \in \Omega$, the mapping $\varphi^\omega : (0, \infty) \times \mathrm{D} \to \mathrm{D}$ is continuous.

The first classical result that we should take into account is that the reflecting Brownian motion $\{Y_t\}_{t\geq 0}$ defines a random dynamical system on D' over the metric dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, (\sigma_t)_{t\geq 0})$. This is a well-known result that, for domains without boundary can be found for instance in [5, Section 2.3] and it is essentially a consequence of

$$B_s(\omega) + B_t(\sigma_s\omega) = \omega(s) + \sigma_s\omega(t) = \omega(t+s) = B_{t+s}(\omega),$$

where $\{\sigma_t\}_{t\geq 0}$ are the standard shift operators on Ω for the Brownian motion, defined as

$$\sigma_t\omega(s) = \omega(s+t) - \omega(s) \quad s, t \geq 0, \omega \in \Omega.$$

With this in mind we define the *skew-product* of the reflecting Brownian motion $\{Y_t\}_{t\geq 0}$ and the shift operators. This is the following mapping,

$$\begin{array}{rccc} \Theta : & (0,\infty) \times \Omega \times \mathrm{D}' & \to & \Omega \times \mathrm{D}' \\ & (t, \omega, y) & \mapsto & (\sigma_t\omega, Y_t^\omega(y)), \end{array}$$

or $\Theta_t(\omega, y) = (\sigma_t\omega, Y_t^\omega(y))$ for short.

Consider the measurable space $(\Omega \times \mathrm{D}, \mathcal{F} \otimes \mathcal{B}(\mathrm{D}'))$ and the probability measure $\mathbb{P} \times \mathrm{Leb}_{\mathrm{D}'}$ on $(\Omega \times \mathrm{D}', \mathcal{F} \otimes \mathcal{B}(\mathrm{D}'))$, where $\mathrm{Leb}_{\mathrm{D}'}$ is the normalized Lebesgue measure on D'. Then we have the following result.

Lemma 5.1. $(\Omega \times \mathrm{D}', \mathcal{F} \otimes \mathcal{B}(\mathrm{D}'), \mathbb{P} \times \mathrm{Leb}_{\mathrm{D}'}, \{\Theta_t\}_{t\geq 0})$ *defines a metric dynamical system.*

Proof. Let us first check that $\{\Theta_t\}_{t\geq 0}$ has the semigroup property. We have that $\Theta_0(\omega, y) = (\sigma_0\omega, Y_0^\omega(y)) = (\omega, y)$ for every $\omega \in \Omega$ and $y \in \mathrm{D}'$. In addition,

$$\begin{aligned} \Theta_t(\Theta_s(\omega, y)) &= \Theta_t(\sigma_s\omega, Y_s^\omega(y)) = (\sigma_t\sigma_s\omega, Y_t^{\sigma_s\omega}(Y_s^\omega(y))) \\ &= (\sigma_{t+s}\omega, Y_{t+s}^\omega(y)) = \Theta_{t+s}(\omega, y) \end{aligned}$$

holds for every $\omega \in \Omega$ and $y \in \mathrm{D}'$, therefore $\{\Theta_t\}_{t\geq 0}$ defines a semigroup on $\Omega \times \mathrm{D}'$.

It remains thus to check that $\mathbb{P} \times \mathrm{Leb}_{\mathrm{D}'}$ is invariant with respect to Θ_t, namely

$$(\Theta_t)_\#(\mathbb{P} \times \mathrm{Leb}_{\mathrm{D}'}) = \mathbb{P} \times \mathrm{Leb}_{\mathrm{D}'} \quad \forall t \geq 0,$$

or equivalently

$$(\sigma_t)_\#\mathbb{P} = \mathbb{P} \quad \text{and} \quad \mathbb{E}[(X_t)_\#\mathrm{Leb}_{\mathrm{D}'}] = \mathrm{Leb}_{\mathrm{D}'} \quad \forall t \geq 0.$$

Note that $\mathbb{P}$ is invariant with respect to σ_t by construction of the Brownian motion. We can prove that the Lebesgue measure is invariant under the action of the Brownian motion since for every $f : \mathrm{D}' \to \mathbb{R}$ smooth enough with Neumann homogeneous boundary conditions there holds

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\int_{\mathrm{D}'} f(Y_t(y))\,\mathrm{d}y = \mathbb{E}\int_{\mathrm{D}'}\left(\nabla f(y)\cdot\frac{\mathrm{d}Y_t(y)}{\mathrm{d}t} + \Delta f\right)\mathrm{d}y = \int_{\mathrm{D}} \nabla f \cdot \mathbb{E}\frac{\mathrm{d}Y_t(y)}{\mathrm{d}t}\,\mathrm{d}y = 0,$$

where we made use of Itô formula and the fact that the expectation of the (reflecting) white noise $\mathrm{d}Y_t/\,\mathrm{d}t$ is zero. ■

With those tools we can prove that the flow $\{X_t\}_{t\geq 0}$ itself forms a random dynamical system. However it cannot be over the Brownian motion itself or the shift operators alone, so we need to redefine the vector field as follows,

$$u(X_t, Y_t) = \tilde{u}(X_t, \Theta_t),$$

where the dependence with Y_t is thus encoded in Θ_t. With this notation we arrive to the main purpose of the section with the following result.

Lemma 5.2. *The flow* $\{X_t\}_{t\geq 0}$ *given by* (5.12) *forms a random dynamical system over the metric dynamical system* $(\Omega \times \mathrm{D}', \mathcal{F} \otimes \mathcal{B}(\mathrm{D}'), \mathbb{P} \times \mathrm{Leb}_{\mathrm{D}'}, \{\Theta_t\}_{t\geq 0})$.

Proof. It suffices to show the that the flow satisfies the cocycle property over Θ_t and we will argue by uniqueness of the ODE (5.12). Let $\omega \in \Omega$, $x \in \mathrm{D}$, $y \in \mathrm{D}'$ and $s \geq 0$. Consider the random functions

$$\xi(t) = X_{t+s}^{\omega,y}(x), \quad \eta(t) = X_t^{\Theta_s(\omega,y)}(X_s^{\omega,y}(x)).$$

Those functions satisfy the same initial condition $\xi(0) = \eta(0) = X_s^{\omega,y}(x)$, and in addition, by the definition of the flow (5.12),

$$\xi'(t) = \partial_t X_{t+s}^{\omega,y}(x) = \tilde{u}(X_{t+s}^{\omega,y}(x), \Theta_{t+s}(\omega, y)) = \tilde{u}(\xi(t), \Theta_{t+s}(\omega, y))),$$

$$\begin{aligned}\eta'(t) &= \partial_t X_t^{\Theta_s(\omega,y)}(X_s^{\omega,y}(x)) = \tilde{u}(X_t^{\Theta_s(\omega,y)}(X_s^{\omega,y}(x)), \Theta_t(\Theta_s(\omega, y))) \\ &= \tilde{u}(\eta(t), \Theta_t(\Theta_s(\omega, y))).\end{aligned}$$

Then by the semigroup property of the the skew-product, $\Theta_{t+s} = \Theta_t \circ \Theta_s$, we get that $\eta(t)$ and $\xi(t)$ satisfy the same ODE with the same initial data and thus by uniqueness of solutions we get that both functions coincide for all $t \geq 0$. ■

Remark 5.1. When it comes to define the RDS, instead of considering the position of the point vortex we could analogously have thought about a suitable Hilbert space $\mathbb{V}$ of vector fields of the form $u = u(x, y)$ where the coordinate y is of the form Y_t^ω and initially $y = Y_0$. Then we could represent the evolution of such vector field by

$$\begin{aligned} U: \quad (0,\infty) \times \Omega \times \mathbb{V} &\to \quad \mathbb{V} \\ (t, \omega, u_0) \quad &\mapsto \quad U_t^\omega(u_0) \end{aligned}$$

and show that this forms a random dynamical system over a suitable metric dynamical system. However, since we use a specific structure for the stochastic process $\{Y_t\}_{t\geq 0}$ (i.e. the Brownian motion) we decided to define $\{\Theta_t\}_{t\geq 0}$ instead.

Once we have a well defined random dynamical system we can apply the vast theory about these objects, namely Furstenberg-Kesten Theorem, Multiplicative Ergodic Theorem,... to obtain positivity of the top Lyapunov exponent,

$$\lambda_+(x, \omega) = \lim_{t\to\infty} \frac{1}{t} \log |\nabla_x X_t^\omega| > 0.$$

This, together with the geometric ergodicity of the Markov processes before mentioned is what yields the sought exponential mixing by the vector field u.

The Lagrangian perspective has been proved to be effective and useful when it comes to deal with type of problems, however in this work we want to introduce a different perspective, where we do not follow the particle trajectories but look at the overall picture of the Eulerian perspective.

5.3. Eulerian perspective

The point of view that we want to introduce about the ergodicity and mixing problem is the so called *Eulerian* perspective, that puts a focus on the PDE. As explained before, any flow of the form

$$\mathrm{d}X_t = u(t, X_t)\,\mathrm{d}t, \quad X_0 = \mathrm{id},$$

has an analogous PDE respresentation in form of transport equation

$$\partial_t \theta + u \cdot \nabla\theta = 0, \quad \theta(0, \cdot) = \theta_0.$$

These are related via the method of characteristics, or Feynman-Kac formula in case the vector field is stochastic.

We want to study analogous properties to what it has been done in the literature though the Markov processes P_t, $\hat{P}_t$ or $P_t^{(2)}$ but using the passive scalar θ. We can translate the language of Markov process to passive scalars and solutions to the PDE via

$$P_t(x,\cdot) = \mathbb{E}\theta(t,\cdot).$$

The other Markov processes introduced in the previous Section are more involved to study and we will give details in a few lines.

In order to proceed with this framework it is certainly very important to consider the specific structure of the stochastic process $\{Y_t\}_{t\geq 0}$, namely the source of our randomness, so that we can take advantage of it in order to obtain information about the passive scalar θ. In particular we consider the flow $\{X_t\}_{t\geq 0}$ coupled with $\{Y_t\}_{t\geq 0}$,

$$\begin{cases} \mathrm{d}X_t &= u(X_t, Y_t)\,\mathrm{d}t & \text{in } \mathrm{D}, \\ \mathrm{d}Y_t &= \sqrt{2\nu}\,\mathrm{d}B_t - n(Y_y)\,\mathrm{d}L_t & \text{in } \mathrm{D}', \end{cases} \tag{5.13}$$

with $(X_0, Y_0) = \mathrm{id}_{\mathrm{D}\times\mathrm{D}'}$. Then we can construct, via Feynman-Kac or backwards Kolmogorov, see Section A.1 in the Appendix, an equivalent partial differential equation where we study the evolution of a passive scalar $f : (0,\infty)\times \mathrm{D}\times \mathrm{D}' \to \mathbb{R}$. This PDE has the form

$$\begin{cases} \partial_t f + u\cdot\nabla_x f &= \nu\Delta_y f & \text{in } (0,\infty)\times\mathrm{D}\times\mathrm{D}', \\ n_y\cdot\nabla_y f &= 0 & \text{on } (0,\infty)\times\mathrm{D}\times\partial\mathrm{D}', \\ f(0,\cdot,\cdot) &= f_0 & \text{in } \mathrm{D}\times\mathrm{D}'. \end{cases} \tag{5.14}$$

Here we assume $\nu > 0$ to be the diffusion coefficient in the y direction and n_y the outer unit vector to the boundary of $\partial\mathrm{D}'$. The vector field $u = u(x,y)$ is the same as for the transport problem (5.1), however, now we consider a new variable $y \in \mathrm{D}'$ in the second entry, and we do not evaluate any stochastic process. This implies in particular that the vector field in (5.14) is autonomous and deterministic.

Therefore, thanks to the specific structure for the randomness in the stochastic transport problem (5.1) in D, we can write a new PDE that is deterministic in the product space $\mathrm{D}\times\mathrm{D}'$ where $x \in \mathrm{D}$ and $y \in \mathrm{D}'$. This is indeed a deterministic advection-diffusion equation, with the advection in the direction of x and the diffusion in the direction of y.

Our objective here is then to obtain similar results to what it is known through the Langrangian lens regarding ergodicity and mixing with random vector fields [7, 8, 12], but now using purely analytical and PDE methods.

The important results in this business concern the decay of solutions to (5.14) in some suitable norm, thus a reader that is familiar with equations like (5.14) can directly realize that this is a problem related to the concepts of *hypocoercivity* and *hypoellipticity.*

These are concepts that are intrinsically related to parabolic equations such as the heat equation or the advection-difussion equation with a divergence-free vector fields. Let us give a general picture in a more abstract setting. We have an evolution PDE in terms of an operator $\mathcal{L}$ as follows,

$$\partial_t h + \mathcal{L}h = 0, \quad h(0,\cdot) = h_0, \tag{5.15}$$

posed in some domain D. If the domain is bounded one might need to consider some appropriate boundary conditions if needed. Then operator $\mathcal{L}$ is defined on a Hilbert space to itself, e.g. $\mathcal{L} : L^2 \to L^2$.

To showcase a clarifying example let us consider an operator $\mathcal{L} : \mathcal{H} \to \mathcal{H}$ where the Hilbert space $\mathcal{H}$ consists of the L^2 function in D with mean zero,

$$\mathcal{H} = \left\{ f \in L^2(\mathrm{D}) : \fint_{\mathrm{D}} f(x)dx = 0 \right\}.$$

Then a simple but enlightening example to consider is $\mathcal{L} = -\Delta$, namely (5.15) is the heat equation. Therefore we know that if we consider $h_0 \in \mathcal{H}$, then $h(t) \in \mathcal{H}$ for all $t > 0$ and moreover $\|h(t)\|_{L^2}$ decays instantaneously and with an exponential rate to zero. When an operator $\mathcal{L}$ has this property we say that is *coercive*.

There are certain operators for which the Hilbert norm does not decay instantaneously but it does maybe after some finite time $T > 0$. Thus, their long time behaviour is as of a coercive operator even though technically they might not be because for short times there might be other dominating phenomena. For these type of operator we have the following concept.

Definition 5.3. Let $\mathcal{H}$ be a Hilbert space and consider an operator $\mathcal{L} : \mathcal{H} \to \mathcal{H}$. If there exist constants $\lambda, C > 0$ such that solutions to the equation (5.15) satisfy

$$\|h(t)\|_{\mathcal{H}} \leq Ce^{-\lambda t}$$

for all $t > 0$, then $\mathcal{L}$ is called *hypocoercive*.

Going back to our specific PDE, notice that in (5.14) we have the operator

$$\mathcal{L}_1 = u \cdot \nabla_x - \nu \Delta_y$$

in $\mathrm{D} \times \mathrm{D}'$. Notice that since $\nabla_x \cdot u = 0$, if we had the full diffusion $\Delta_x + \Delta_y$ in $\mathrm{D} \times \mathrm{D}'$ then the situation would be totally analogous to the heat equation and the operator $\mathcal{L}_1$ would be coercive. Since we do not have this property we need to work a bit harder to prove hypocoercivity.

Due to the structure of our specific problem, we want to consider a initial datum for (5.14) of the form

$$f(0,x,y) = \theta_0(x)\rho(y), \quad x \in \mathrm{D}, \quad y \in \mathrm{D}'.$$

Here θ_0 is the initial datum of the original passive scalar problem (5.1), and ρ is the initial law for the (reflecting) Brownian motion driving the vector field. We will assume that

$$\theta_0 \in (H^1 \cap L^\infty)(\mathrm{D}), \quad ⨍_{\mathrm{D}} \theta_0(x)dx = 0, \quad \rho \in (W^{1,\infty} \cap \mathcal{P})(\mathrm{D}').$$

Therefore the operator $\mathcal{L}_1$ is defined on the Hilbert space of H^1 functions in $\mathrm{D} \times \mathrm{D}'$ with mean zero. That is a good definition for the operator $\mathcal{L}_1$ since,

- θ_0 mean zero and $\rho \geq 0$ imply that f_0 and thus $f(t)$ is mean zero for all $t > 0$;
- $\theta_0 \in H^1(\mathrm{D})$ and $\rho \in W^{1,\infty}(\mathrm{D}')$ imply that $f_0 \in H^1(\mathrm{D} \times \mathrm{D}')$;
- $f_0 \in H^1(\mathrm{D} \times \mathrm{D}')$ implies that $f(t) \in H^1(\mathrm{D} \times \mathrm{D}')$ for all $t > 0$.

Remark 5.2 (Integrability of ρ)*.* Regarding the initial law $\rho \in \mathcal{P}(\mathrm{D}')$ for the process $\{Y_t\}_{t\geq 0}$, we could in principle consider any Radon probability measure, including atomic measures. Since Y_t is a Brownian motion, its law solves a heat equation and thus any initial datum ρ will be instantaneously smooth. With a singular measure ρ we could rewrite all the statements of this chapter in terms a new initial time $t_0 > 0$. We just choose $\rho \in W^{1,\infty}(\mathrm{D}')$ in order to avoid this technicality.

At this point we find a double objective: first to study whether hypocoercivity of the operator $\mathcal{L}_1$ implies some of the ergodicity or mixing properties that we know from the Langrangian perspective, and second find examples of vector fields for which we get hypocoercivity.

For the first part we find that studying solutions to (5.14), namely the operator $\mathcal{L}_1$, will only be enough to obtain ergodicity. This is to be expected since $f(t)$ is constructed from solutions to the stochastic passive scalar θ that is in turn related to the one-point Markov process $P_t = \mathbb{E}\theta(t,\cdot)$. For this reason and in orther to be consisten with the already existing literature, we will refer to the operator $\mathcal{L}_1$ as the *one-point operator*, or the operator associated to the *one-point problem.* Properties of this operator and examples of vector fields u for which $\mathcal{L}_1$ is hypocoercive are provided in Section 5.4.

5.3.1. The hypocoercivity method

In [95], Villani designs an abstract framework to prove that a operators like $\mathcal{L}_1$ are hypocoercive. Given some Hilbert space $\mathcal{H}$ that is L^2-based and some natural number

$d \geq 1$ consider operators

$$A : \mathcal{H} \to \mathcal{H}^d, \quad B : \mathcal{H} \to \mathcal{H}.$$

If we denote $A^\star$ the adjoint operator of A then it satisfies $A^\star : \mathcal{H}^d \to \mathcal{H}$. We assume that B is antisymmetric, namely $B^\star = -B$, and define the operator

$$\mathcal{L} = A^\star A + B.$$

In $\mathbb{R}^d$, a typical example would be to consider A to be a gradient, and thus $A^\star$ (minus) a divergence, which yields a diffusion operator $\mathcal{L} = -\Delta$. If in addition we include $B = u \cdot \nabla$ with u divergence free, then we obtain an advection-diffusion operator $\mathcal{L} = u \cdot \nabla - \Delta$.

Our specific one-point and two-point operators fall under this abstract structure as well, since $A_{\mathcal{L}_1} = A_{\mathcal{L}_2} = \sqrt{\nu}\nabla_y$ have the corresponding gradient structure, and the drift operators are antisymmetric,

$$B_{\mathcal{L}_1} = u \cdot \nabla_x, \quad B_{\mathcal{L}_2} = u_1 \cdot \nabla_{x_1} + u_2 \cdot \nabla_{x_2}$$

because of the divergence-free property of the vector field.

Since it was introduced in [95], the method has been used in many different type of problems, for instance to obtain enhanced dissipation with similar operators to $\mathcal{L}_1$, see e.g. [10, 28]. Before proceeding with the method, let us clarify some notation. We assume the operator $\mathcal{L}$ is defined in a L^2-based Hilbert space $\mathcal{H}$, that usually can be understood as some sort of weighted H^1 space.

- Let $\|\cdot\|$ be the standard L^2 norm and $\|\cdot\|_{H^1}$ the standard H^1 norm, namely

$$\|h\|_{H^1}^2 = \|h\|^2 + \|\nabla h\|^2.$$

- Let $\langle \cdot, \cdot \rangle$ be the standard scalar product in L^2, namely $\langle h, h \rangle = \|h\|^2$.

Villani's method consists of defining a new norm, called *augmented energy functional* Φ, that is equivalent to the original norm for which one wants to prove the exponential decay. The advantage of using Φ is that the operator $\mathcal{L}$ will be coercive with that norm, even if it is not with the original norm of $\mathcal{H}$. Thus, when applied to solutions to the equation $\partial_t h + \mathcal{L}h = 0$, it decays instantaneously.

The $\mathcal{H}$ norm and the augmented energy functional typically involve first order derivatives, therefore we say that this method provides a frame to deal with H^1-hypocoercivity. That is one of the main reasons for us to consider initial data $\theta_0, f_0, F_0 \in H^1$.

Villani designs his augmented energy functional in terms of some commutators from the operators A and B. The idea is to consider a first commutator

$$C_1 h = [A, B]h = ABh - BAh \in \mathcal{H}^d \quad \text{for all } h \in \mathcal{H},$$

and then proceed iteratively

$$C_{k+1} = [C_k, B], \quad k \geq 1.$$

We want to consider as many iterations as needed to obtain a norm that is comparable to the standard H^1 norm. In this respect let us introduce a more refined definition of coercivity, that is equivalent to the concept mentioned before.

Definition 5.4. We say that an operator $C : \mathcal{H} \to \mathcal{H}$ is *coercive* if there exists a constant $\lambda > 0$ such that

$$\langle h, Ch \rangle \geq \lambda \|h\|^2$$

for all $h \in \mathcal{H}$.

With this information in mind we can proceed to explain the structure of the $\mathcal{H}$ norm and the augmented energy functional. Consider the collection of the $N + 1$ first commutators $\{C_i : 0 \leq i \leq N\}$, where we denote $C_0 = A$.

- We define $\| \cdot \|_{\mathcal{H}}$ to be the norm of the Hilbert space $\mathcal{H}$ via the commutators

$$\|h\|_{\mathcal{H}}^2 = \|h\|^2 + \sum_{i=0}^{N} \|C_i h\|^2.$$

- We define the augmented energy functional $\Phi(\cdot)$ as

$$\Phi(h) = \|h\|^2 + \sum_{i=0}^{N} (\alpha_i \|C_i h\|^2 + 2\beta_i \langle C_i h, C_{i+1} \rangle),$$

  where $\alpha_i, \beta_i > 0$ for all $0 \leq i \leq N$ and $C_{N+1} = 0$.

What Villani proved in [95, Theorem 24] is that under certain assumptions for the commutators $\{C_i : 0 \leq i \leq N\}$, there exist appropriate constants $\alpha_i, \beta_i > 0$ for which $\| \cdot \|_{\mathcal{H}}$ is equivalent to $\Phi(\cdot)$. In addition, if the operator

$$\sum_{i=0}^{N} C_i^{\star} C_i \quad \text{is coercive,} \tag{5.16}$$

then we obtain that $\mathcal{L}$ is hypocoercive with the $\mathcal{H}$ norm.

The main ideas about this result can be understood from the following steps:

1. For any h that satisfies $\partial_t h + \mathcal{L}h = 0$ we prove that there exists $\varepsilon > 0$ such that

$$\frac{d}{dt}\Phi(h(t)) + \varepsilon \sum_{i=0}^{N} \|C_i h\|^2 \leq 0$$

2. Condition (5.16) means that

$$\sum_{i=0}^{N} \langle h, C_i^\star C_i h\rangle = \sum_{i=0}^{N} \|C_i h\|^2 \gtrsim \|h\|^2,$$

   and thus we can write,

$$\frac{d}{dt}\Phi(h(t)) + \varepsilon \|h\|_{\mathcal{H}}^2 \lesssim 0.$$

   Since $\Phi(\cdot)$ and $\|\cdot\|_{\mathcal{H}}^2$ are equivalent, we use Gronwall's inequality to obtain the instantaneous and exponential decay $\Phi(h(t)) \leq \Phi(h(0))e^{-\varepsilon t}$.

3. Yet again, via the equivalence between $\Phi(\cdot)$ and $\|\cdot\|_{\mathcal{H}}^2$ we obtain the decay of the latter, $\|h(t)\|_{\mathcal{H}}^2 \lesssim \|h_0\|_{\mathcal{H}}^2 e^{-\varepsilon t}$, which might not be instantaneous for this norm but still exponential.

So far this proves that $\mathcal{L}$ is $\mathcal{H}-$hypocoercive. However, for some specific cases, we can use the particular structure of the commutators $C_1, \ldots, C_N$ to recover the standard H^1 norm from the $\mathcal{H}$ norm. This is something that can be easily understood from a particular example.

Example 5.1. Consider the one-point operator associated to equation (5.14),

$$\mathcal{L}_1 = u \cdot \nabla_x - \nu \Delta_y, \quad A = \sqrt{\nu}\nabla_y, \quad B = u \cdot \nabla_x.$$

The zeroth commutator $C_0 = A$ already provides half of the derivatives needed for the standard H^1 norm in $\mathrm{D} \times \mathrm{D}'$. However, if we want to construct from $\mathcal{L}_1$ a norm that is comparable to the standard H^1 norm, we need to recover the missing derivatives from the commutators. Notice that all of them are of the form

$$C_k = [C_{k-1}, B] = M_k \nabla_x, \quad M_k \in \mathbb{R}^{d\times d}, \quad k \geq 1,$$

e.g. $M_1 = \nabla_y u^T$, $M_2 = \nabla_y u^T \nabla_x u^T - (u \cdot \nabla_x)\nabla_y u^T$,... For this particular example we define the $\mathcal{H}$ norm,

$$\|h\|_{\mathcal{H}}^2 = \|h\|^2 + \|\nabla_y h\|^2 + \sum_{i=1}^{N} \|M_i \nabla_x h\|^2.$$

One can appreciate that the number N of commutators needed is related to the minimum number of factors $M_i \nabla_x$ required to span $\mathbb{R}^d$, see Remark 5.3. In addition, if we have a suitable ellipticity property for the collection of matrices M_i like

$$\sum_{i=1}^{N} \|M_i \xi\|^2 \gtrsim \|\xi\|^2, \quad \text{for all } \xi \in \mathbb{R}^d,$$

then we can recover the original H^1 norm from $\mathcal{H}$, and we obtain hypocoercivity in the usual H^1 sense.

Remark 5.3 (Relation to Hörmander's hypoellipticity). Conditions to obtain Villani's hypocoercivity resemble very much conditions to obtain Hörmander's hypoellipticity. In [53], Hörmander considers similar operators

$$\mathcal{L} = \sum_{i=1}^{N} A_i^\star A_i + A_0.$$

He proves that if the operator $\mathcal{L}$ has some specific structure, then $\mathcal{L}$ is hypoelliptic. Hypoellipticity in this sense is related to the smoothing properties of the operator $\mathcal{L}$, namely $\mathcal{L}$ is said to be *hypoelliptic* if any solution to $\partial_t h + \mathcal{L} h = f$ with f smooth is smooth, regardless of the regularity of the initial datum. *Hörmander's condition* for the operator $\mathcal{L}$ to be hypocoelliptic is the following:

- the collection of commutators

  $$[A_i, A_j], [[A_i, A_j], A_k], [[[A_i, A_j], A_k], A_l], \ldots$$

  for all $i, j, k, l, \ldots \in \{0, \ldots, N\}$, span the the full space $\mathbb{R}^M$ where the equation is defined.

Thus, one can see the resemblance between Hörmander's conditions and the coercivity of the operator $C_1^\star C_1 + \ldots C_N^\star C_N$ from Villani's Theorem.

In Sections 5.5 and 5.6 we use the hypocoercivity method to prove the H^1 decays of the solutions to the equation 5.14 with different examples of vector fields. The augmented energy functional that we introduce is heavily inspired by Villani's, but with some suitable changes that are more convenient for our specific situation. We do not apply Villani's Theorem directly, but we follow similar arguments with a similar augmented energy functional to obtain our result.

5.4. Ergodicity results: the one-point process

The main purpose of this section is to prove Theorem 5.1 that in the coming sections will be used to provide examples of vector fields that will produce exponential ergodicity for solutions to the transport equation (5.1) in two different types of domains.

On the one hand we consider the 2-dimensional torus $\mathbb{T}^2$ and shear flows with random phase, which is an example already known in the literature to produce mixing, see e.g. [12]. The novelty here is the method, since we follow a Eulerian perspective in order to show ergodicity and annealed mixing. On the other hand we consider the domain D to be the unit ball in $\mathbb{R}^2$ and D' to be a ball contained in D with smaller radius. This is new example since it features a domain with boundary and therefore the derivation of hypocoercivity if far more involved.

The main assumption in order to obtain the sought ergodicity result is the exponential decay of solution to a certain PDE. This particular equation is denoted as the *one-point problem* (5.5) since in the Lagrangian perspective it features the dynamics of a one particle trajectory $X_t \in \mathrm{D}$ starting at $x \in \mathrm{D}$. The name is also originated from the literature, see [6, 7, 8, 12], where the authors define a Markov process

$$P_t(x, A) = \mathbb{P}[X_t \in A \mid X_0 = x]$$

that defines the trajectory of one particle (or *one-point*). From the PDE perspective this Markov process corresponds to the expected value of density solution to the transport problem (5.1), namely

$$P_t(x, \cdot) = \mathbb{E}\theta(t, x).$$

In this context we say that the solution $\theta(t, \cdot)$ to (5.1) is ergodic if the associated Markov process $P_t = \mathbb{E}\theta(t, \cdot)$ decays in time to its average. This decay can occur pointwise in $x \in \mathrm{D}$ (see e.g. [12]) or in some suitable norm. Since for the transport problem we consider our solution to be in H^1, it makes sense for us to consider a decay of the L^2 in order to measure ergodicity. This motivates the following definition.

Definition 5.5. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a domain $\mathrm{D} \subset \mathbb{R}^d$, we say that a time dependent function $g = g(t, x, \omega)$ is *ergodic* if

$$\left\| \mathbb{E}g(t, \cdot, \cdot) - \mathbb{E}\fint_{\mathrm{D}} g(t, x, \cdot)\,\mathrm{d}x \right\|_{L^2(\mathrm{D})} \to 0 \quad \text{as } t \to \infty.$$

In particular, we say g is *exponentially ergodic* is there is some constant $\lambda > 0$ such that

$$\left\| \mathbb{E}g(t, \cdot, \cdot) - \mathbb{E}\fint_{\mathrm{D}} g(t, x, \cdot)\,\mathrm{d}x \right\|_{L^2(\mathrm{D})} \lesssim e^{-\lambda t}$$

for all $t > 0$.

For our particular problem we will always consider initial data that is mean zero for all $t \geq 0$ and all $\omega \in \Omega$, and therefore we look for results of the type,

$$\|\mathbb{E}g(t, \cdot, \cdot)\|_{L^2(\mathrm{D})} \lesssim e^{-\lambda t}.$$

This definition might look slightly surprising when applied to solutions to the transport equation (5.1), since the L^2 is conserved

$$\|\theta(t, \cdot)\|_{L^2(\mathrm{D})} = \|\theta_0\|_{L^2(\mathrm{D})} \quad \text{for all } t \geq 0.$$

However this conservation law is not such when dealing with the expectation of $\theta(t, \cdot)$. In order to see that we can integrate the transport equation in Ω,

$$\partial_t \mathbb{E}\theta + \mathbb{E}[u \cdot \nabla\theta] = 0.$$

Then, testing with $\mathbb{E}\theta(t, \cdot)$ and integrating in D we obtain

$$\frac{1}{2}\frac{d}{dt}\|\mathbb{E}\theta(t, \cdot)\|^2_{L^2(\mathrm{D})} + \int_{\mathrm{D}} \mathbb{E}[u \cdot \nabla\theta(t, x)]\mathbb{E}\theta(t, x)\,\mathrm{d}x = 0,$$

and therefore one sees that unless $u \cdot \nabla\theta$ and θ would be independent for all $t > 0$ (which cannot be), the norm $\|\mathbb{E}\theta(t, \cdot)\|_{L^2}$ is not conserved in time.

Without further ado, we proceed with the proof of Theorem 5.1. For this result we are assuming that the solution f to the extended problem (5.5) decays exponentially and then we prove that this yields exponential ergodicity for the marginal θ, namely the solution to (5.1). The challenging point for the examples given later will be to prove the decay of the solutions to the extended problem f.

Proof of Theorem 5.1. We are interested in the time evolution of the $L^2(\mathrm{D})$ norm of $\mathbb{E}\theta$, that we will define by duality

$$\|\mathbb{E}\theta(t,\cdot)\|_{L^2}^2 = \sup_{\|g\|_{L^2(\mathrm{D})}=1} \int_{\mathrm{D}} \mathbb{E}\theta(t,x)g(x)\,\mathrm{d}x.$$

Then, take some function $g \in L^2(\mathrm{D})$ with $\|g\|_{L^2(\mathrm{D})} = 1$ so that we can write

$$\int_{\mathrm{D}} \mathbb{E}\theta(t,x)g(x)\,\mathrm{d}x = \mathbb{E}\int_{\mathrm{D}} g(x)(X_t)_{\#}[\theta_0(x)]\,\mathrm{d}x = \mathbb{E}\int_{\mathrm{D}} g(X_t(x))\theta_0(x)\,\mathrm{d}x.$$

Now take $\rho \in \mathcal{P}(\mathrm{D}')$ and the stochastic process $\{Y_t\}_{t\geq 0}$ with law $(Y_0) = \rho$. We integrate everything in D' with a new variable that we denote by $y \in \mathrm{D}'$. Then we have that

$$\begin{aligned}\int_{\mathrm{D}} \mathbb{E}\theta(t,x)g(x)\,\mathrm{d}x &= \mathbb{E}\iint_{\mathrm{D}\times\mathrm{D}'} g(X_t(x))\theta_0(x)\rho(y)\,\mathrm{d}y \\ &= \iint_{\mathrm{D}\times\mathrm{D}'} g(x)\mathbb{E}\left[(X_t,Y_t)_{\#}(\theta_0\otimes\rho)\right](x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \iint_{\mathrm{D}\times\mathrm{D}'} g(x)f(t,x,y)\,\mathrm{d}x\,\mathrm{d}y,\end{aligned}$$

where we have used that $f(t,x,y) = \mathbb{E}\left[(X_t,Y_t)_{\#}(\theta_0\otimes\rho)\right](x,y)$. The this function $f(t,x,y)$ is the solution to the one-point problem (5.5), since $\{(X_t,Y_t)\}_{t\geq 0}$ is the coupled stochastic process (5.4). Now, directly by a Hölder argument we arrive to

$$\int_{\mathrm{D}} \mathbb{E}\theta(t,x)g(x)\,\mathrm{d}x \leq \|f(t,\cdot,\cdot)\|_{L^2(\mathrm{D}\times\mathrm{D}')}\|g\|_{L^2(\mathrm{D})},$$

but by assumption $\|f(t,\cdot,\cdot)\|_{L^2(\mathrm{D}\times\mathrm{D}')}$ decays exponentially fast, so there yields

$$\int_{\mathrm{D}} \mathbb{E}\theta(t,x)g(x)\,\mathrm{d}x \lesssim e^{-\lambda t}$$

for all $g \in L^2(\mathrm{D})$ with $\|g\|_{L^2(\mathrm{D})} = 1$. ■

Some straightforward consequence of the ergodicity is the so called annealed mixing, or mixing averaged on the noise. We define it as follows.

Definition 5.6. Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and $U : (0,\infty)\times\mathrm{D}\times\Omega \to \mathbb{R}^d$ a random vector field. We say that U produces *annealed mixing* if the solution to the transport equation $\theta(t,\cdot)$ with vector field U and initial datum $\theta_0 \in H^1(\mathrm{D})$ mean zero satisfies

$$\|\mathbb{E}\theta(t,\cdot)\|_{H^{-1}(\mathrm{D})} \to 0 \quad \text{as } t\to\infty.$$

Therefore it is easy to see that if $\theta(t,\cdot)$ is ergodic, then there is also annealed mixing in the sense of Definition 5.6 since

$$\|\mathbb{E}\theta(t,\cdot)\|_{H^{-1}(\mathrm{D})} = \sup_{\|\phi\|_{H^1}=1} \int_{\mathrm{D}} \mathbb{E}\theta(t,x)\phi(x)\,\mathrm{d}x \lesssim \|\mathbb{E}\theta(t,\cdot)\|_{L^2(\mathrm{D})}.$$

In the next sections we will focus on finding examples of random vector fields in different types of domains for which the H^1 norm of solutions to the one-point problem (5.5) decays exponentially fast. The example here presented are the following:

1. First we consider a bounded domain without boundary, $\mathrm{D} = \mathbb{T}^2$ and $\mathrm{D}' = \mathbb{T}^2$. Here we will define shear flows with a random phase as in [12, 27], but not alternating since we need a continuous in time version of the problem.

2. The second example concerns a domain with boundary, namely $\mathrm{D} = B_1(0)$. In this domain we consider a smooth version of a point vortex that moves randomly in a subdomain $\mathrm{D}' = B_r(0)$ with $r \in (0,1)$. The random path that the point vortex follows is a reflecting Brownian motion, and we obtain exponential ergodicity for any $r > 0$.

We will do so using the method of hypocoercivity in $\mathrm{D} \times \mathrm{D}'$ that we explained in the previous section, but we will derive the explicit coercive augmented energy functional that yields the corresponding exponential decay. Therefore, for the two examples here presented, the hypocoercivity result plus Theorem 5.1 yield that solutions to the transport equations are exponentially ergodic in the sense of Definition 5.5.

5.5. Example 1: Shear flows with random phase

The first example of mixing vector field that we want to consider is the case of a shear flow in $\mathbb{T}^2$ with random phase. Following the notation presented in the previous section, we define a vector field of the form

$$u(x,y) = \begin{pmatrix} \sin(x_2 + y_1) \\ \sin(x_1 + y_2) \end{pmatrix}, \tag{5.17}$$

where $x = (x_1, x_2) \in \mathrm{D} = \mathbb{T}^2$ and $y = (y_1, y_2) \in \mathrm{D}' = \mathbb{T}^2$.

We define the stochastic process $(Y_t)_{t\geq 0}$ to be an standard Brownian motion in $\mathbb{T}^2$ and evaluate u on its second component with $Y_t \in \mathbb{T}^2$. Then we obtain a non-autonomous random vector field $U(t,x) = u(x, Y_t)$ on $\mathbb{T}^2$ and we are in the setting to apply all the theory we introduce in the previous sections.

As before, consider a mean zero initial datum $\theta_0 \in H^1(\mathbb{T}^2)$ and an initial law $\rho \in W^{1,\infty}(\mathbb{T}^2)$ for the Brownian motion. We define the following deterministic PDE in $\mathbb{T}^2 \times \mathbb{T}^2$,

$$\begin{cases} \partial_t f + u(x,y) \cdot \nabla_x f &= \nu \Delta_y f \quad \text{in } (0,\infty) \times \mathbb{T}^2 \times \mathbb{T}^2, \\ f(0,\cdot) &= \theta_0 \otimes \rho \quad \text{in } \mathbb{T}^2 \times \mathbb{T}^2. \end{cases} \tag{5.18}$$

The idea behind this theory is to prove that the $H^1(\mathbb{T}^2 \times \mathbb{T}^2)$ norm of solutions $f(t,\cdot,\cdot)$ to this equation decay exponentially fast in time to $f \equiv 0$. Thanks to Theorem **??**, this then yields that the random vector field $U(t,x)$ mixes exponentially fast $\theta_0 \in H^1(\mathbb{T}^2)$ almost-surely.

In order to address the decay of the $H^1(\mathbb{T}^2 \times \mathbb{T}^2)$ norm what we do is studying hypocoercivity of the operator

$$\mathcal{L}_1 = u \cdot \nabla_x - \nu \Delta_y$$

in $\mathbb{T}^2 \times \mathbb{T}^2$. First of all, let us state some of the properties of this vector field, that will be useful in the hypocoercivity estimate. A straightforward computation yields that

$$\nabla_y u(x,y) = \begin{pmatrix} \cos(x_2+y_1) & 0 \\ 0 & \cos(x_1+y_2) \end{pmatrix},$$

and therefore there holds

$$\det \nabla_y u(x,y) = \cos(x_2+y_1)\cos(x_1+y_2),$$

and the matrix norm is

$$|\nabla_y u(x,y)|^2 = \operatorname{tr}(\nabla_y u^T \nabla_y u)(x,y) = \cos^2(x_2+y_1) + \cos^2(x_1+y_2).$$

Then the matrix $\nabla_y u$ has the following property,

$$|\nabla_y u^T \xi|^2 = \xi \cdot \nabla_y u \nabla_y u^T \xi \geq \frac{\cos^2(x_2+y_1)\cos^2(x_1+y_2)}{\cos^2(x_2+y_1)+\cos^2(x_1+y_2)} |\xi|^2$$

for all $\xi \in \mathbb{R}^2$. In particular,

$$|\nabla_y u^T \xi|^2 \geq \frac{\cos^2(x_2+y_1)\cos^2(x_1+y_2)}{2} |\xi|^2 = \frac{(\det \nabla_y u(x,y))^2}{2} |\xi|^2$$

for all $\xi \in \mathbb{R}^2$.

In light of this estimate for the $y-$derivatives of the vector field, we prove the following

Hardy-type estimate.

Lemma 5.3 (Hardy estimate)**.** *Let $u(x,y)$ be the vector field from* (5.17)*. Then given any $g=g(y)$, $g\in H^1(\mathbb{T}^2)$, there holds*

$$\|g\|^2_{L^2(\mathbb{T}^2)} \lesssim \|\det(\nabla_y u(x,\cdot))g\|^2_{L^2(\mathbb{T}^2)} + \|\nabla_y g\|^2_{L^2(\mathbb{T}^2)}$$

for all $x\in\mathbb{T}^2$.

Proof. Let us fix $x\in\mathbb{T}^2$, by translation invariance in x we can consider without loss of generality $x=0$, therefore the estimate we want to prove is

$$\int_{\mathbb{T}^2}|g|^2\,\mathrm{d}(y_1,y_2) \lesssim \int_{\mathbb{T}^2}(\cos(y_1)\cos(y_2))^2|g|^2\,\mathrm{d}(y_1,y_2) + \int_{\mathbb{T}^2}|\nabla_y g|^2\,\mathrm{d}(y_1,y_2).$$

First of all let us address the $1-$dimensional problem $y\in\mathbb{T}=[0,2\pi)$. We define $\eta:\mathbb{T}\to[0,1]$ smooth and such that

$$\eta(y)=\begin{cases} 1 & \text{if } y\in\left[\frac{\pi}{4},\frac{3\pi}{4}\right]\cup\left[\frac{5\pi}{4},\frac{7\pi}{4}\right], \\ 0 & \text{if } y\in\left[0,\frac{\pi}{8}\right]\cup\left[\frac{7\pi}{8},\frac{9\pi}{8}\right]\cup\left[\frac{15\pi}{8},2\pi\right). \end{cases}$$

Then, we can write

$$\int_{\mathbb{T}}|g|^2\,\mathrm{d}y = \int_{\mathbb{T}}(1-\eta)|g|^2\,\mathrm{d}y + \int_{\mathbb{T}}\eta|g|^2\,\mathrm{d}y.$$

The first addend is straightforwardly controlled by

$$\int_{\mathbb{T}}(1-\eta)|g|^2\,\mathrm{d}y \lesssim \int_{\mathbb{T}}(1-\eta)(\cos(y))^2|g|^2\,\mathrm{d}y \le \int_{\mathbb{T}}(\cos(y))^2|g|^2\,\mathrm{d}y.$$

For the second addend we need to make use the splitting

$$\int_{\mathbb{T}}\eta|g|^2\,\mathrm{d}y \lesssim \int_0^{\pi}\eta\sin(y)|g|^2\,\mathrm{d}y - \int_{\pi}^{2\pi}\eta\sin(y)|g|^2\,\mathrm{d}y,$$

then, by means of integration by parts and Young's inequality,

$$\begin{aligned}\int_0^\pi \eta|g|^2\,\mathrm{d}y &\lesssim \int_0^\pi \eta\sin(y)|g|^2\,\mathrm{d}y = \int_0^\pi \eta'\cos(y)|g|^2\,\mathrm{d}y + 2\int_0^\pi \eta\cos(y)gg'\,\mathrm{d}y \\ &\lesssim \int_0^\pi |\cos(y)||g|^2\,\mathrm{d}y + \int_0^\pi |\cos(y)||g||g'|\,\mathrm{d}y \\ &\leq \int_0^\pi |\cos(y)|^2|g|^2\,\mathrm{d}y + \frac{1}{4}\int_0^\pi |g|^2 + \frac{1}{2}\int_0^\pi |\cos(y)|^2|g|^2\,\mathrm{d}y + \frac{1}{2}\int_0^\pi |g'|^2\,\mathrm{d}y,\end{aligned}$$

and analogously for the remaining integral in $[\pi, 2\pi)$. All in all, we arrive to the $1-$dimensional estimate

$$\|g\|^2_{L^2(\mathbb{T})} \lesssim \|\cos(\cdot)g\|^2_{L^2(\mathbb{T})} + \|g'\|^2_{L^2(\mathbb{T})}.$$

For the $2-$dimensional case then we can write

$$\int_{\mathbb{T}^2} |g(y_1,y_2)|^2\,\mathrm{d}(y_1,y_2) = \int_{\mathbb{T}} G(y_2)\,\mathrm{d}y_2$$

where, thanks to the estimate for the $1-$dimensional problem we have

$$G(y_2) = \int_{\mathbb{T}} |g(y_1,y_2)|^2\,\mathrm{d}y_1 \lesssim \int_{\mathbb{T}} (\cos(y_1))^2|g(y_1,y_2)|^2\,\mathrm{d}y_1 + \int_{\mathbb{T}} \left|\frac{\partial g(y_1,y_2)}{\partial y_1}\right|^2\mathrm{d}y_1.$$

Therefore, by an analogous procedure we can also write

$$\int_{\mathbb{T}^2} |g(y_1,y_2)|^2\,\mathrm{d}(y_1,y_2) \lesssim \int_{\mathbb{T}} (\cos(y_1))^2 H(y_1)\,\mathrm{d}y_1 + \int_{\mathbb{T}^2} \left|\frac{\partial g(y_1,y_2)}{\partial y_1}\right|^2\mathrm{d}(y_1,y_2)$$

where, again due to the $1-$ dimensional estimate we obtain

$$H(y_1) = \int_{\mathbb{T}} |g(y_1,y_2)|^2\,\mathrm{d}y_2 \lesssim \int_{\mathbb{T}} (\cos(y_2))^2|g(y_1,y_2)|^2\,\mathrm{d}y_2 + \int_{\mathbb{T}} \left|\frac{\partial g(y_1,y_2)}{\partial y_2}\right|^2\mathrm{d}y_2.$$

Gathering everything together

$$\begin{aligned}\int_{\mathbb{T}^2} |g(y_1,y_2)|^2\,\mathrm{d}(y_1,y_2) &\lesssim \int_{\mathbb{T}^2} (\cos(y_1))^2(\cos(y_2))^2|g(y_1,y_2)|^2\,\mathrm{d}(y_1,y_2) \\ &\quad + \int_{\mathbb{T}^2}\left(\left|\frac{\partial g(y_1,y_2)}{\partial y_1}\right|^2 + (\cos(y_1))^2\left|\frac{\partial g(y_1,y_2)}{\partial y_2}\right|^2\right)\mathrm{d}(y_1,y_2),\end{aligned}$$

which implies the desired result. ■

Remark 5.4. An important consequence of this Hardy estimate is that, if we integrate in the x variable too and use the properties of the vector field before mentioned, we obtain the estimate

$$\|g\|^2_{L^2(\mathbb{T}^2\times\mathbb{T}^2)} \lesssim \|\nabla_y u^T g\|^2_{L^2(\mathbb{T}^2\times\mathbb{T}^2)} + \|\nabla_y g\|^2_{L^2(\mathbb{T}^2\times\mathbb{T}^2)} \tag{5.19}$$

for all $g \in H^1(\mathbb{T}^2 \times \mathbb{T}^2)$.

With these tools in hand now we can face the hypocoercivity problem for the operator $\mathcal{L}_1 = u\cdot\nabla_x - \nu\Delta_y$ with u given by (5.17). From now on and for the sake of a clearer notation, we will denote the norm in $L^2(\mathbb{T}^2\times\mathbb{T}^2)$ simply by $\|\cdot\|$. In addition, the standard scalar product in $L^2(\mathbb{T}^2\times\mathbb{T}^2)$ will be denoted by $\langle\cdot,\cdot\rangle$.

Lemma 5.4. *Let f be a solution to* (5.18) *and consider the augmented energy functional*

$$\Phi(f) = \frac{1}{2}\|f\|^2 + \frac{\alpha}{2}\|\nabla_y f\|^2 + \beta\langle\nabla_y f, \nabla_y u^T\nabla_x f\rangle + \frac{\gamma}{2}\|\nabla_x f\|^2 \tag{5.20}$$

with $\alpha,\beta,\gamma > 0$. Then there holds

$$\begin{aligned}\frac{d}{dt}\Phi(f) = &-\nu\|\nabla_y f\|^2 - \alpha\nu\|\Delta_y f\|^2 - \beta\|\nabla_y u^T\nabla_x f\|^2 - \gamma\nu\|\nabla^2_{xy} f\|^2\\ &- \alpha\langle\nabla_y f, \nabla_y u^T\nabla_x f\rangle - 2\beta\nu\langle\Delta_y f, \nabla_y(u\cdot\nabla_x)\cdot\nabla_y f\rangle\\ &- \beta\nu\langle\Delta_y f, \Delta_y u\cdot\nabla_x f\rangle - \beta\langle u\cdot\nabla_x\nabla_y f, \nabla_y u^T\nabla_x f\rangle\\ &+ \beta\langle\nabla_y(u\cdot\nabla_x)\cdot\nabla_y f, u\cdot\nabla_x f\rangle - \gamma\langle\nabla_x f, \nabla_x u^T\nabla_x f\rangle.\end{aligned}$$

Proof. In order to obtain this result we must consider $\partial_t f = -u\cdot\nabla_x f + \nu\Delta_y f$ and perform the appropriate integrations by parts. First of all we have

$$\frac{1}{2}\frac{d}{dt}\|f\|^2 = \langle f, \partial_t f\rangle = -\langle f, u\cdot\nabla_x f\rangle + \nu\langle f, \Delta_y f\rangle = -\nu\|\nabla_y f\|^2.$$

For the α terms we can write

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\|\nabla_y f\|^2 &= \langle\nabla_y f, \nabla_y\partial_t f\rangle = -\langle\nabla_y f, \nabla_y(u\cdot\nabla_x f)\rangle + \nu\langle\nabla_y f, \nabla_y\Delta_y f\rangle\\ &= -\langle\nabla_y f, \nabla_y u^T\nabla_x f\rangle - \langle\nabla_y f, (u\cdot\nabla_x)\nabla_y f\rangle - \nu\|\Delta_y f\|^2\\ &= -\langle\nabla_y f, \nabla_y u^T\nabla_x f\rangle - \nu\|\Delta_y f\|^2.\end{aligned}$$

Analogously, for the γ terms we have

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\|\nabla_x f\|^2 &= \langle \nabla_x f, \nabla_x \partial_t f\rangle = -\langle \nabla_x f, \nabla_x (u\cdot\nabla_x f)\rangle + \nu\langle \nabla_x f, \nabla_x \Delta_y f\rangle \\ &= -\langle \nabla_x f, \nabla_x u^T \nabla_x f\rangle - \langle \nabla_x f, (u\cdot\nabla_x)\nabla_x f\rangle - \nu\|\nabla^2_{xy} f\|^2 \\ &= -\langle \nabla_x f, \nabla_x u^T \nabla_x f\rangle - \nu\|\nabla^2_{xy} f\|^2.\end{aligned}$$

Finally, the crossed-terms or β terms can be computed via

$$\begin{aligned}\frac{d}{dt}\langle \nabla_y f, \nabla_y u^T \nabla_x f\rangle &= \langle \nabla_y \partial_t f, \nabla_y u^T \nabla_x f\rangle + \langle \nabla_y f, \nabla_y u^T \nabla_x \partial_t f\rangle \\ &= -\langle \nabla_y (u\cdot\nabla_x f), \nabla_y u^T \nabla_x f\rangle + \nu\langle \nabla_y \Delta_y f, \nabla_y u^T \nabla_x f\rangle \\ &\quad - \langle \nabla_y f, \nabla_y u^T \nabla_x (u\cdot\nabla_x f)\rangle + \nu\langle \nabla_y f, \nabla_y u^T \nabla_x \Delta_y f\rangle \\ &= -\|\nabla_y u^T \nabla_x f\|^2 - \langle u\cdot\nabla_x \nabla_y f, \nabla_y u^T \nabla_x f\rangle - \nu\langle \Delta_y f, \Delta_y u\cdot\nabla_x f\rangle \\ &\quad - \nu\langle \Delta_y f, \nabla_y (u\cdot\nabla_x)\cdot\nabla_y f\rangle + \langle \nabla_y (u\cdot\nabla_x)\cdot\nabla_y f, u\cdot\nabla_x f\rangle \\ &\quad - \nu\langle \nabla_y (u\cdot\nabla_x)\cdot\nabla_y f, \Delta_y f\rangle.\end{aligned}$$

Therefore, combining all four computations we arrive to the claim of the Lemma. ■

Lemma 5.5. *Let f be a solution to (5.18) and $\Phi(f)$ be the augmented energy functional from (5.20). There exists a choice of $\alpha, \beta, \gamma > 0$ such that:*

1. *$\Phi(\cdot)$ is equivalent to $\|\cdot\|^2_{H^1}$ as a norm in $\mathbb{T}^2\times\mathbb{T}^2$, namely*

$$\min\left\{1, \frac{1}{\nu^2}\right\}\|f\|^2_{H^1} \lesssim \Phi(f) \lesssim \max\left\{1, \frac{1}{\nu^2}\right\}\|f\|^2_{H^1}.$$

2. *If $\nu > 1$ is sufficiently large, then there exist $\delta_1, \delta_2 > 0$ independent of ν so that there holds*

$$\frac{d}{dt}\Phi(f) + \frac{\nu}{2}\|\nabla_y f\|^2 + \frac{\delta_1}{\nu}\|\nabla_y u^T \nabla_x f\|^2 + \frac{\delta_2}{\nu}\|\nabla^2_{xy} f\|^2 \le 0$$

for all $t > 0$.

Proof. For the first point, recall the definition of the augmented energy functional (5.20),

$$\Phi(f) = \frac{1}{2}\|f\|^2 + \frac{\alpha}{2}\|\nabla_y f\|^2 + \beta\langle \nabla_y f, \nabla_y u^T \nabla_x f\rangle + \frac{\gamma}{2}\|\nabla_x f\|^2.$$

Notice that since $\|\nabla_y u\|_\infty = 1$, via Hölder and Young's inequalities we can write

$$\beta|\langle \nabla_y f, \nabla_y u^T \nabla_x f\rangle| \leq \frac{\beta}{2}\sqrt{\frac{\alpha}{\gamma}}\|\nabla_y f\|^2 + \frac{\beta}{2}\sqrt{\frac{\gamma}{\alpha}}\|\nabla_x f\|^2,$$

which in turn implies

$$\Phi(f) \geq \frac{1}{2}\|f\|^2 + \frac{\alpha}{2}\left(1 - \frac{\beta}{\sqrt{\alpha\gamma}}\right)\|\nabla_y f\|^2 + \frac{\gamma}{2}\left(1 - \frac{\beta}{\sqrt{\alpha\gamma}}\right)\|\nabla_x f\|^2,$$

$$\Phi(f) \leq \frac{1}{2}\|f\|^2 + \frac{\alpha}{2}\left(1 + \frac{\beta}{\sqrt{\alpha\gamma}}\right)\|\nabla_y f\|^2 + \frac{\gamma}{2}\left(1 + \frac{\beta}{\sqrt{\alpha\gamma}}\right)\|\nabla_x f\|^2.$$

Therefore $\Phi(\cdot)$ and $\|\cdot\|^2_{H^1}$ will be equivalent norms provided that $0 < \beta < \sqrt{\alpha\gamma}$. If then we choose the coefficients to be of the form

$$\alpha = a, \quad \beta = \frac{b}{\nu}, \quad \gamma = \frac{c}{\nu^2},$$

where $a, b, c > 0$ are independent of ν, then the first claim of the Lemma follows. The reason for this specific choice of coefficients will become apparent within the next lines.

The second statement of the Lemma is a by-product of Hölder and Young's inequalities. We start with the α term,

$$\alpha|\langle \nabla_y f, \nabla_y u^T \nabla_x f\rangle| \leq \frac{\nu}{2}\|\nabla_y f\|^2 + \frac{\alpha^2}{2\nu}\|\nabla_y u^T \nabla_x f\|^2.$$

The γ term has to be treated using Hardy's inequality (5.19). In order to keep track of all the exact coefficients we will use Hardy's inequality in the following form

$$\|\nabla_x f\|^2 \leq k_1\|\nabla_y u^T \nabla_x f\|^2 + k_2\|\nabla^2_{xy} f\|^2.$$

Then we can write

$$\gamma|\langle \nabla_x f, \nabla_x u^T \nabla_x f\rangle| \leq \gamma\|\nabla_x f\|^2 \leq \gamma k_1\|\nabla_y u^T \nabla_x f\|^2 + \gamma k_2\|\nabla^2_{xy} f\|^2,$$

where we have used the property $\|\nabla_x u\|_\infty = 1$.

We deal with all the β terms in a similar way, making use of Hardy's inequality (5.19) when necessary and taking into account that $\|u\|_\infty, \|\Delta_y u\|_\infty = 1$. Then, we obtain

$$2\beta\nu|\langle \Delta_y f, \nabla_y(u \cdot \nabla_x) \cdot \nabla_y f\rangle| \leq \frac{\alpha\nu}{2}\|\Delta_y f\|^2 + \frac{2\beta^2\nu}{\alpha}\|\nabla^2_{xy} f\|^2,$$

$$\begin{aligned}\beta\nu|\langle\Delta_y f, \Delta_y u\cdot\nabla_x f\rangle| &\leq \frac{\alpha\nu}{4}\|\Delta_y f\|^2 + \frac{\beta^2\nu}{\alpha}\|\nabla_x f\|^2 \\ &\leq \frac{\alpha\nu}{4}\|\Delta_y f\|^2 + \frac{\beta^2\nu k_1}{\alpha}\|\nabla_y u^T\nabla_x f\|^2 + \frac{\beta^2\nu k_2}{\alpha}\|\nabla^2_{xy} f\|^2,\end{aligned}$$

$$\beta|\langle u\cdot\nabla_x\nabla_y f, \nabla_y u^T\nabla_x f\rangle| \leq \beta\|\nabla^2_{xy} f\|^2 + \frac{\beta}{4}\|\nabla_y u^T\nabla_x f\|^2,$$

$$\begin{aligned}\beta|\langle\nabla_y(u\cdot\nabla_x)\cdot\nabla_y f, u\cdot\nabla_x f\rangle| &\leq \beta\|\nabla^2_{xy} f\|^2 + \beta\|\nabla^2_{xy} f\|\|\nabla_y u^T\nabla_x f\| \\ &\leq 2\beta\|\nabla^2_{xy} f\|^2 + \frac{\beta}{4}\|\nabla_y u^T\nabla_x f\|^2.\end{aligned}$$

Introducing these bounds in the result from Lemma 5.4 we arrive to the stability estimate

$$\frac{d}{dt}\Phi(f) + \frac{\nu}{2}\|\nabla_y f\|^2 + \frac{\alpha\nu}{4}\|\Delta_y f\|^2 + \mathrm{I}(\nu)\|\nabla_y u^T\nabla_x f\|^2 + \mathrm{II}(\nu)\|\nabla^2_{xy} f\|^2 \leq 0,$$

where

$$\mathrm{I}(\nu) = \frac{\beta}{2} - \frac{\alpha^2 k_1}{2\nu} - \gamma k_1 - \frac{\beta^2\nu}{\alpha},$$

$$\mathrm{II}(\nu) = (\nu - k_2)\gamma - \frac{(2+k_2)\beta^2\nu}{\alpha} - 3\beta.$$

Now we need to make sure that $\mathrm{I}(\nu), \mathrm{II}(\nu) > 0$, but first notice that if $\mathrm{II}(\nu) > 0$ then there holds

$$\beta^2 < \alpha\gamma,$$

which is the necessary (and sufficient) condition for Φ to be equivalent to the standard H^1 norm

$$\|f\|^2_{H^1} = \|f\|^2 + \|\nabla_y f\|^2 + \|\nabla_x f\|^2.$$

One reasonable ansatz we can make in order to find the appropriate coefficients $\alpha, \beta, \gamma > 0$ is

$$\alpha = a, \quad \beta = \frac{b}{\nu}, \quad \gamma = \frac{c}{\nu^2},$$

where $a, b, c > 0$ are independent of ν. Then the compatibility conditions that we obtain are the following. On the one hand, for $\mathrm{I}(\nu) > 0$ we need

$$\frac{b}{2\nu} > \frac{a^2 k_1}{2\nu} + \frac{b^2}{A\nu} + \frac{ck_1}{\nu^2},$$

that can be achieved by choosing

$$\frac{b}{2} > \frac{a^2k_1}{2} + \frac{b^2}{a} \quad \text{and } \nu > 1 \text{ sufficiently large.}$$

On the other hand, for $\mathrm{II}(\nu) > 0$ we need

$$\frac{c}{\nu} - \frac{ck_2}{\nu^2} > \frac{(2+k_2)b^2}{a\nu} + \frac{3b}{\nu},$$

that can be obtained by choosing

$$c > \frac{(2+k_2)b^2}{a} + 3b \quad \text{and } \nu > 1 \text{ sufficiently large.}$$

Therefore we obtain a compatibility region for the coefficients in terms of $k_1, k_2 > 0$ of the form

$$a \in \left(0, \frac{1}{8k_1}\right), \quad b \in \left(\frac{a}{4}(1-\sqrt{1-8ak_1}), \frac{a}{4}(1+\sqrt{1-8ak_1})\right), \quad c > \frac{(2+k_2)b^2}{a} + 3b.$$

All in all, taking into account that $\mathrm{I}(\nu)$ scales like β and $\mathrm{II}(\nu)$ scales like $\gamma\nu$, we see that both coefficients are of the form $1/\nu$ and there yields the second claim of the Lemma. ■

Remark 5.5. Notice that Hardy's inequality (5.19) and this Lemma together imply that

$$\frac{d}{dt}\Phi(f) + \nu\|\nabla_y f\|^2 + \frac{1}{\nu}\|\nabla_x f\|^2 \lesssim 0,$$

where the constant absorbed in $\lesssim$ does not depend on $\nu > 1$. Then, since f has mean zero, we can apply Poincaré inequality to obtain

$$\frac{d}{dt}\Phi(f) + \frac{1}{\nu}\|f\|^2 + \nu\|\nabla_y f\|^2 + \frac{1}{\nu}\|\nabla_x f\|^2 \lesssim 0,$$

which in turn implies

$$\frac{d}{dt}\Phi(f) + \frac{1}{\nu}\|f\|^2_{H^1(\mathbb{T}^2\times\mathbb{T}^2)} \lesssim 0.$$

All these partial results can be gathered together to obtain full hypocoercivity of the operator $\mathcal{L}_1 = u \cdot \nabla_x - \nu\Delta_y$. This is the main statement of the following Proposition.

Proposition 5.1. *Let* $f(t,\cdot)$ *be a solution to* (5.18) *in* $\mathbb{T}^2 \times \mathbb{T}^2$ *and assume that* $\nu > 1$ *is* *sufficiently large. Then there holds*

$$\|f(t,\cdot)\|_{H^1} \lesssim \nu \|f(0,\cdot)\|_{H^1} e^{-\frac{1}{2\nu}t}$$

for all $t > 0$ *and where the constants absorbed in* $\lesssim$ *do not depend on* ν*.*

Proof. This result is an straightforward application of all the partial results studied before. On the one hand we know that

$$\frac{d}{dt}\Phi(f) + \frac{1}{\nu}\|f\|^2_{H^1(\mathbb{T}^2\times\mathbb{T}^2)} \lesssim 0.$$

On the other hand we have

$$\frac{1}{\nu^2}\|f\|^2_{H^1} \lesssim \Phi(f) \lesssim \|f\|^2_{H^1}.$$

Then we can write

$$\frac{d}{dt}\Phi(f) + \frac{1}{\nu^2}\Phi(f) \lesssim 0,$$

so a combination of Gronwall's inequality and the equivalence between $\Phi(\cdot)$ and $\|\cdot\|^2_{H^1}$ implies the claim of the Proposition. ∎

5.6. Example 2: Randomly moving vortex on a disc

The next example considered deals with a vector field on a bounded domain with boundary. In particular we define $\mathrm{D} = B_1(0) \subset \mathbb{R}^2$ the ball of radius 1. In this domain we consider the stream function ψ induced by a vortex located at $y \in \mathrm{D}$, that is,

$$\begin{cases} -\Delta\psi(\cdot,y) &= \delta_y \quad \text{in D}, \\ \psi &= 0 \quad \text{on } \partial\mathrm{D}. \end{cases} \tag{5.21}$$

Given any fixed value of y, the Poisson problem (5.21) has an explicit solution given by

$$\psi(x,y) = \frac{1}{2\pi}\log\left(\frac{|y||x-y^*|}{|x-y|}\right), \quad \text{where } y^* = \frac{y}{|y|^2}.$$

Notice that, using the definition of $y^\star$, the term in the numerator can be rewritten as

$$|y||x-y^*| = \sqrt{1 - 2x\cdot y + |x|^2|y|^2}.$$

This stream function gives rise to the vector field generated by a singular point vortex located at y via the perpendicular gradient, namely

$$u_{\mathrm{PV}}(x,y) = \nabla_x^{\perp}\psi(x,y),$$

which is an unbounded and singular vector field with $u_{\mathrm{PV}}(\cdot,y) \in L^p(\mathrm{D})$ if and only if $1 \leq p < 2$, and therefore $\nabla_x u_{\mathrm{PV}}$ is not even integrable.

We want to consider a regularized version of this velocity field, so we define

$$u(x,y) = -\frac{d(x,y)^2}{1-|y|^2}\nabla_x^{\perp}\psi(x,y),$$

where

$$d(x,y) = e^{-2\pi\psi(x,y)}$$

approximates the Euclidean distance between x and y in a neighborhood of the vortex y, and regularizes thus the singular velocity field $\nabla_x^{\perp}\psi$ induced by the point vortex. The y dependent prefactor is just introduced for convenience as it will slightly simplify our later computations. If y is essentially located near the origin, this term is of order one and has no substantial effect on the velocity. In order to avoid problems when $|y|$ is very close to 1 we will define the domain of y to be $\mathrm{D}' = B_r(0)$ with $r \in (0,1)$.

All in all, the vector field that we consider for this example takes the following form

$$u(x,y) = \frac{y^{\perp}(1-|x|^2) - x^{\perp}(1-|x|^2+|x-y|^2)}{2\pi(1-2x\cdot y+|x|^2|y|^2)^2}. \tag{5.22}$$

Observe that there are a couple of particular cases which are interesting to be studied. For instance, from (5.22) one can see that when the center of the vortex lies in the center of the domain, i.e. $y=0$, then this vector field is nothing but a rigid rotation,

$$u(x,0) = -\frac{1}{2\pi}x^{\perp}.$$

In addition, it is also remarkable that as expected, one can obtain from (5.22) that the vector field is always tangential to the boundary $\partial\mathrm{D}$,

$$u(x,y)\Big|_{x\in\partial\mathrm{D}} = -\frac{x^{\perp}|x-y|^2}{2\pi(1-2x\cdot y+|x|^2|y|^2)^2}.$$

As it has been pointed out, it is evident that if we consider a vortex lying at $y=0$ and we let it rest there, no mixing can be achieved since this vector field produce nothing but a rigid rotation. The main idea behind this vector field comes thus from considering such smooth vortex with center at $y \in \mathrm{D}'$ and letting the center moved around the domain $\mathrm{D}' \subset \mathrm{D}$. We do not want to consider *any* possible trajectory but a reflecting Brownian motion, hence this example will be referred in the future pages as *randomly moving vortex*. In this situation the goal is to study the mixing and ergodic properties of u in the disc D.

More in detail and recalling the notation introduced before, we want to consider the stochastic process

$$dY_t = \sqrt{2\nu} dB_t - n(Y_t) dL_t$$

in $(0, \infty) \times \mathrm{D}'$ with $Y_0 = \mathrm{id}$, where $\{B_t\}_{t \geq 0}$ is a standard Brownian motion in $\mathbb{R}^2$, $n(\cdot)$ is the outer normal unit vector to the boundary $\partial \mathrm{D}'$ and $(L_t)_{t \geq 0}$ is a local time of the process $(Y_t)_{t \geq 0}$ that is activated if and only if the Brownian path touches the boundary $\partial \mathrm{D}'$. In such case the path is reflected when touching the boundary according to the normal direction n to $\partial \mathrm{D}'$. At the level of the PDE, this ensures Neumann homogeneous boundary conditions.

With the vector field $u(x, Y_t)$ we consider the transport equation

$$\begin{cases} \partial_t \theta + u(\cdot, Y_t) \cdot \nabla \theta &= 0 \quad \text{in } (0, \infty) \times \mathrm{D}, \\ \theta(0, \cdot) &= 0 \quad \text{in } \mathrm{D}, \end{cases}$$

that produces a solution θ that depends on the noise realization $\omega \in \Omega$ of u. As before, we want to study whether the L^2 norm of $\mathbb{E}\theta$ decays exponentially in time. In order to do so we define the extended problem

$$\begin{cases} \partial_t f + u \cdot \nabla_x f &= \nu \Delta_y f \quad \text{in } (0, \infty) \times \mathrm{D} \times \mathrm{D}', \\ n_y \cdot \nabla_y f &= 0 \quad \text{on } (0, \infty) \times \mathrm{D} \times \partial \mathrm{D}', \\ f(0, \cdot, \cdot) &= \theta_0 \otimes \rho \quad \text{in } \mathrm{D} \times \mathrm{D}', \end{cases} \tag{5.23}$$

and the main goal of this section will be to prove that the H^1 norm of the solution to (5.23) decays exponentially in time. This will yield ergodicity of the vector field $u(\cdot, Y_t)$ in D as proved in Theorem 5.1.

One straightforward outcome of the the definition for this vector field is that it has the following properties:

1. For any fixed $y \in \mathrm{D}'$, the vector field $u(x, y)$ is tangential to $\partial \mathrm{D}$. This is a byproduct of the fact that the vector field generated by the singular point vortex u_{PV} has the same property, since the streamlines of u and u_{PV} are in the same position in the domain D.

2. The vector field $u(x,y)$ is divergence free in x, namely

$$\nabla_x \cdot u(x,y) = 0 \quad \text{for all} \quad y \in \mathrm{D}'.$$

Notice that if we write $u(x,y) = V(\psi(x,y),y)\nabla_x^\perp \psi(x,y)$, then a direct computation yields

$$\nabla_x \cdot u = V(\psi,y)\nabla_x \cdot \nabla_x^\perp \psi + \partial_1 V(\psi,y)\nabla_x \psi \cdot \nabla_x^\perp \psi = 0.$$

Just for reference, we would like to highlight that u induces a smooth vorticity distribution in D, that is,

$$\omega(x,y) = \nabla_x \times u(x,y) = -\frac{1}{\pi}\frac{1-|y|^2}{\sqrt{1-2x\cdot y+|x|^2|y|^2}}.$$

Notice that it will never blow up provided that $|y| \le r < 1$. In addition one can observe that if $|y| = 0$ there is a uniform vorticity distribution in the whole domain D,

$$\omega(x,0) = -\frac{1}{\pi},$$

which corresponds, as stated before, to a rigid motion of the particles in the the domain around the origin.

In view of the argument that we follow to prove hypocoercivity, we want to know when $\nabla_y u(x,y)$ is degenerating. Recall from the previous sections that, for any $A \in \mathbb{R}^{2\times 2}$ we have the estimate

$$\xi \cdot A^T A\xi = |A\xi|^2 \ge \frac{(\det A)^2}{\operatorname{tr}(A^T A)}|\xi|^2 \quad \text{for all } \xi \in \mathbb{R}^2.$$

Therefore we are interested in the determinant of $\nabla_y u$,

$$\det \nabla_y u(x,y) = \frac{1-|x|^2}{4\pi}\frac{1+2x\cdot y-3|x|^2}{(1-2x\cdot y+|x|^2|y|^2)^4}.$$

The y gradient of the velocity field degenerates on two lines: on the one hand it degenerates on the $x-$boundary $|x| = 1$, and on the other hand it degenerates on the circle

$$\left|x - \frac{1}{3}y\right|^2 = \frac{|y|^2}{9} + \frac{1}{3},$$

which lies in the interior of the domain $\mathrm{D} = B_1(0)$ for any interior (fixed) point y.

Remark 5.6 (The degeneracy at $|x| = 1$). On a first look, one can hint why $|x| = 1$ presents

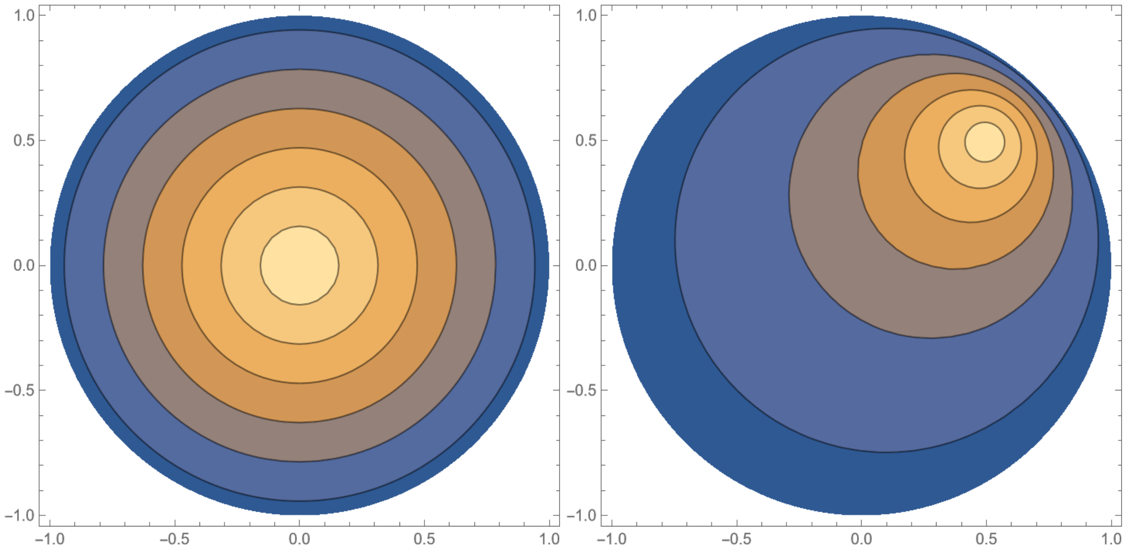

Figure 5.1.: Contour lines of the stream function ψ when $y = (0, 0)$ and $y = (0.5, 0.5)$.

a degeneracy. The vector field u is always tangential to $\partial\mathrm{D}$ by definition, therefore no matter how we change the position of the center of the vortex y, there will always be one direction for which points with $|x| = 1$ will not change their velocity: the normal direction to the boundary, namely,

$$\frac{\partial}{\partial y_i} u(x, y) \cdot \frac{x}{|x|} = 0 \quad \text{for all } x \in \partial\mathrm{D} \text{ and } i \in \{1, 2\}.$$

Notice that this degeneracy is not sharp, if we choose any other direction $\xi \in \mathbb{R}^2$ not normal to the boundary, i.e. $\xi \cdot x^\perp \neq 0$, we will get that $\nabla_y u^T \xi \neq 0$ if $|x| = 1$. In particular this means that the large eigenvalue of the matrix $\nabla_y u(x, y)$ does not degenerate when $|x| = 1$.

Remark 5.7 (The ring degeneracy). The understanding of the circle degeneracy around $y \in \mathrm{D}'$ is certainly more involved. First of all, it is remarkable that if we consider the singular point vortex u_{PV}, this degeneracy does not occur. In such case, only when $|x| = 1$ the determinant of $\nabla_y u_{\mathrm{PV}}$ vanishes for the same reason that is happens to (5.22). However, as soon as we consider a smooth version of the point vortex (and there are several ways to do so), the ring degeneracy pops up.

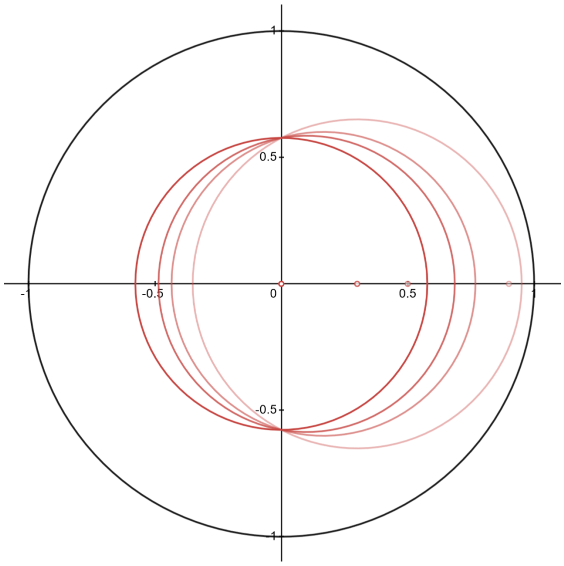

Figure 5.2.: Sketch of the ring where the degeneracy occurs for $y = (0, 0)$, $y = (0.3, 0)$, $y = (0.5, 0)$ and $y = (0.9, 0)$ from left to right.

One can spot, directly from the definition

$$\left| x - \frac{1}{3} y \right|^2 = \frac{|y|^2}{9} + \frac{1}{3},$$

or from the sketch in Figure 5.2, that the ring will be of smallest radius when $y = 0$. Then, as $y \to \partial \mathrm{D}$, the radius grows but always stays completely inside D as long as $|y| < 1$, that we are imposing by choosing $r < 1$. It is noticeable as well that if y moves radially from the origin of the domain, there are exactly two points that will remain in the ring as y moves.

We want to deal with these degeneracies independently, so first of all let us work with the ring. From the $y \in \mathrm{D}'$ perspective, for any $x \in \mathrm{D}$ fixed, the degeneracy consists of a line, namely

$$1 + 2x \cdot y - 3|x|^2 = 0,$$

that we denote by

$$\ell(x) = \mathrm{D}' \cap \left\{ y = \left(\frac{3}{2} - \frac{1}{2|x|^2} \right) x + s x^{\perp} : s \in \mathbb{R} \right\}.$$

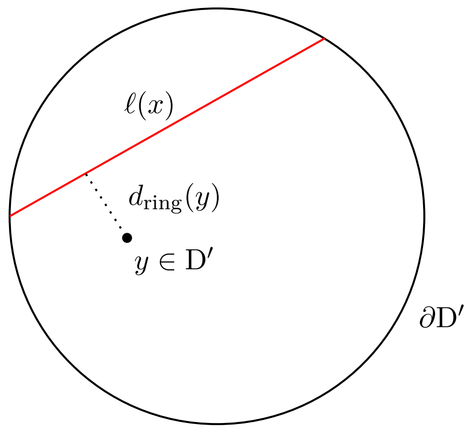

Figure 5.3.: Sketch of the degeneracy line in D' for a fixed value of $x \in \mathrm{D}$.

We define then the following distance for any coordinate $y \in \mathrm{D}'$ to the line $\ell(x)$ as

$$d_{\text{ring}}(y) = \text{dist}(y, \ell(x) \cap \partial \mathrm{D}')$$

and we will use this function plus a Poincaré-Hardy type of estimate in order to deal with this degeneracy of $\nabla_y u^T$. On a first comment, notice that

$$d_{\text{ring}}(y) \leq d_{\mathrm{D}'}(y) = \text{dist}(y, \partial \mathrm{D}') = \frac{r^2 - |y|^2}{2r},$$

where we are assuming that $\mathrm{D}' = B_r(0) \subset \mathbb{R}^2$, for some $r \in (0,1)$.

Lemma 5.6. *Fix $x \in \mathrm{D}$ and let $d_{\text{ring}}(y)$ be defined as before, then there holds*

$$\|g\|_{L^2(\mathrm{D}')} \lesssim \|d_{\text{ring}}^{3/2} g\|_{L^2(\mathrm{D}')} + \|\nabla_y g\|_{L^2(\mathrm{D}')}$$

for all $g = g(y) \in H^1(\mathrm{D}')$.

Proof. First of all let us fix $x \in \mathrm{D}$, $x \neq 0$. In case $x = 0$ then the line is just a constant $\ell(0) = 1$ and $\text{dist}(y, \ell(0)) > 0$ since $y \in B_r(0)$ and $r < 1$.

Let $\varepsilon > 0$ be small enough. In order to prove this result we need to introduce some smooth cut-off function $\eta_\varepsilon : \mathrm{D}' \to [0,1]$ of the form

$$\eta_\varepsilon(y) = \begin{cases} 1 & \text{if } d_{\text{ring}}(y) < \varepsilon, \\ 0 & \text{if } d_{\text{ring}}(y) > 2\varepsilon, \end{cases}$$

and such that

$$\|\nabla_y \eta_\varepsilon\|_{L^\infty} \lesssim \frac{1}{\varepsilon}.$$

Therefore we can write

$$\int_{D'} (1-\eta_\varepsilon) g^2 dy \leq \frac{1}{\varepsilon^3} \int_{D'} (1-\eta_\varepsilon) d_{\text{ring}}^3 g^2 dy \leq \frac{1}{\varepsilon^3} \|d_{\text{ring}}^{3/2} g\|_{L^2}^2.$$

In addition, since $|\nabla_y d_{\text{ring}}(y)| = 1$ for all $y \in D'$ such that $y \notin \ell(x)$, i.e. for almost all $y \in D'$, then via integration by parts we obtain,

$$\begin{aligned}
\int_{D'} \eta_\varepsilon g^2 dy &= \int_{D'} \eta_\varepsilon |\nabla_y d_{\text{ring}}|^2 g^2 dy \\
&= -2 \int_{D'} \eta_\varepsilon g \nabla_y d_{\text{ring}} \cdot \nabla_y g dy - \int_{D'} \nabla_y \eta_\varepsilon \cdot \nabla_y d_{\text{ring}} g^2 d_{\text{ring}} dy \\
&\quad - \int_{D'} \eta_\varepsilon g^2 d_{\text{ring}} \Delta_y d_{\text{ring}} dy \\
&\lesssim \int_{D'} |g||\nabla_y g| dy + \frac{1}{\varepsilon} \int_{D'} g^2 d_{\text{ring}} dy + \int_{D'} g^2 d_{\text{ring}} |\Delta_y d_{\text{ring}}| dy \\
&\lesssim \varepsilon \int_{D'} g^2 dy + \frac{1}{\varepsilon} \int_{D'} |\nabla_y g|^2 dy + \frac{1}{\varepsilon^3} \int_{D'} d_{\text{ring}}^2 g^2 dy,
\end{aligned}$$

where we have used Hölder and Young inequalities and the fact that $\Delta_y d_{\text{ring}} \in L^\infty$.

All in all if we write the splitting

$$\int_{D'} g^2 dy = \int_{D'} (1-\eta_\varepsilon) g^2 dy + \int_{D'} \eta_\varepsilon g^2 dy,$$

what we obtain is

$$(1-\varepsilon) \int_{D'} g^2 dy \lesssim \frac{1}{\varepsilon} \int_{D'} |\nabla_y g|^2 dy + \frac{1}{\varepsilon^3} \int_{D'} d_{\text{ring}}^2 g^2 dy.$$

At this point we can also absorb all the $\varepsilon > 0$ in $\lesssim$ since we will not need this parameter later. Therefore we found that,

$$\int_{D'} g^2 dy \lesssim \int_{D'} |\nabla_y g|^2 dy + \int_{D'} d_{\text{ring}}^2 g^2 dy,$$

which is almost the result we are looking for. Now we only use Young's inequality once more in the following way

$$d_{\text{ring}}^2 \leq \frac{1}{3\varepsilon^2} d_{\text{ring}}^3 + \frac{2\varepsilon}{3},$$

which yields

$$\int_{\mathrm{D}'} d_{\mathrm{ring}}^2 g^2 dy \lesssim \frac{1}{\varepsilon^2}\int_{\mathrm{D}'} d_{\mathrm{ring}}^3 g^2 dy + \varepsilon \int_{\mathrm{D}'} g^2 dy$$

and, after the choice of an appropriate ε coefficient, there yields the desired result. ∎

Remark 5.8. Notice that that Lemma 5.6 holds also for vector-valued functions

$$g = (g_1, \dots, g_n) \in \mathbb{R}^n.$$

In such case, the result holds true for each component $g_i \in \mathbb{R}$ and in particular also in the following form

$$\begin{aligned}\|g\|_{L^2(\mathrm{D}')}^2 = \sum_{i=1}^n \|g_i\|_{L^2(\mathrm{D}')}^2 &\lesssim \sum_{i=1}^n \left(\|d_{\mathrm{ring}}^{3/2} g_i\|_{L^2(\mathrm{D}')}^2 + \|\nabla_y g_i\|_{L^2(\mathrm{D}')}^2\right)\\ &= \|d_{\mathrm{ring}}^{3/2} g\|_{L^2(\mathrm{D}')}^2 + \|\nabla_y g\|_{L^2(\mathrm{D}')}^2.\end{aligned}$$

Before proceeding with the relation between this Lemma and structure of $\nabla_y u$, let us address the other type of degeneracy that this matrix has, namely $\det \nabla_y u(x,y) = 0$ if $|x| = 1$ for all $y \in \mathrm{D}'$. In order to do so we will stated a Lemma with some key properties of the vector field u and its relation with the distance function to the boundary in x,

$$d_{\mathrm{D}}(x) = \mathrm{dist}(x, \partial\mathrm{D}) = 1 - |x|^2.$$

In order to do so let us consider a suitable change of variables for this geometry, namely we will use *polar coordinates* in $x = (r, \phi)$. We define the polar coordinates of the vector field by

$$u = u_r \hat{e}_r + u_\phi \hat{e}_\phi,$$

where

$$u_r(x,y) = u(x,y)\cdot\frac{x}{|x|}, \quad u_\phi(x,y) = u(x,y)\cdot\frac{x^\perp}{|x|}$$

and $\hat{e}_r$, $\hat{e}_\phi$ are unit vectors in the radial and angular directions from the origin.

Lemma 5.7. *Let u be the vector field defined before and let $d_{\mathrm{D}}(x) = 1 - |x|^2$ be the distance to the boundary in x. Then u has the following properties:*

1. *The radial component of u satisfies*

$$|u_r|(x,y) \lesssim d_{\mathrm{D}}(x),$$

for all $(x,y) \in \mathrm{D} \times \mathrm{D}'$. *In particular there holds,*

$$|\nabla_y^n u_r|(x,y), |\partial_\phi^n u_r|(x,y) \lesssim d_{\mathrm{D}}(x)$$

for all $n \geq 0$ *and all* $(x,y) \in \mathrm{D} \times \mathrm{D}'$.

2. *The* 2×2 *matrix defined by*

$$M = \nabla_y u^T \nabla_x u^T - (u \cdot \nabla_x) \nabla_y u^T$$

has the following property

$$|M \nabla_x f|(x,y) \lesssim d_{\mathrm{D}}(x) |\partial_r f|(x,y) + \left| \frac{1}{r} \partial_\phi f \right| (x,y)$$

for all $(x,y) \in \mathrm{D} \times \mathrm{D}'$.

Proof. For the first result we only need to compute the radial component of u, namely

$$u(x,y) \cdot x = -\frac{(1-|x|^2)(x \wedge y)}{2\pi(1 - 2x \cdot y + |x|^2|y|^2)},$$

where $x \wedge y = x_1 y_2 - x_2 y_1$. Then just recalling the definition of $u_r = u \cdot x/|x|$ we arrive to the first statement of the Lemma.

In order to see the second result let us treat the two addends separately. On the one hand, using that $|\nabla_y u|, \nabla_x u| \lesssim 1$, we can write

$$\begin{aligned}
|\nabla_y u^T \nabla_x u^T \nabla_x f| &\leq |\nabla_y u^T \nabla_x u_r \partial_r f| + \left| \nabla_y u^T \nabla_x u_\phi \frac{1}{r} \partial_\phi f \right| \\
&\lesssim |\nabla_y u_r \partial_r u_r \partial_r f| + \left| \frac{1}{r} \nabla_y u_\phi \partial_\phi u_r \partial_r f \right| + \left| \frac{1}{r} \partial_\phi f \right| \\
&\lesssim |\nabla_y u_r||\partial_r f| + |\partial_\phi u_r||\partial_r f| + \left| \frac{1}{r} \partial_\phi f \right|.
\end{aligned}$$

Now using the previous observation that $|\nabla_y u_r|, |\partial_\phi u_r| \lesssim d_{\mathrm{D}}(x)$, there yields

$$|\nabla_y u^T \nabla_x u^T \nabla_x f| \lesssim d_{\mathrm{D}} |\partial_r f| + \left| \frac{1}{r} \partial_\phi f \right|.$$

On the other hand for the second addend we can write,

$$\begin{aligned}|u\cdot\nabla_x\nabla_y u^T\nabla_x f| &\le |u_r\partial_r\nabla_y u^T\nabla_x f| + \left|\frac{1}{r}u_\phi\partial_\phi\nabla_y u^T\nabla_x f\right| \\ &\lesssim |u_r||\nabla_x f| + \left|\frac{1}{r}u_\phi\partial_\phi\nabla_y u_r\partial_r f\right| + \left|\frac{1}{r^2}u_\phi\partial_\phi\nabla_y u_\phi\partial_\phi f\right| \\ &\lesssim |u_r|\left(|\partial_r f| + \left|\frac{1}{r}\partial_\phi f\right|\right) + |\partial_\phi\nabla_y u_r||\partial_r f| + \left|\frac{1}{r}\partial_\phi f\right| \\ &\lesssim d_{\mathrm{D}}|\partial_r f| + \left|\frac{1}{r}\partial_\phi f\right|,\end{aligned}$$

and there we arrive to the claim of the Lemma. ■

The matrix M introduced in Lemma 5.7 will be of relevance when proving hypocoercivity since it corresponds to the second commutator in Villani's hypocoercivity method, see [95].

With all these tools in hand we can derive the following Poincaré-Hardy inequality for the vector field u that will be a key element in order to obtain the sought hypocoercivity.

Proposition 5.2 (Poincaré-Hardy)**.** *Let u be the vector field defined above and let $d_{\mathrm{D}'}(y)$ and $d_{\mathrm{D}}(x)$ be the distances to the boundary $\partial\mathrm{D}'$ and $\partial\mathrm{D}$ respectively. Then, there holds*

$$\|d_{\mathrm{D}}\partial_r f\|_{L^2}^2 + \left\|\frac{1}{r}\partial_\phi f\right\|_{L^2}^2 \lesssim \|\sqrt{d_{\mathrm{D}'}}\nabla_y u^T\nabla_x f\|_{L^2}^2 + \|d_{\mathrm{D}}\partial_r\nabla_y f\|_{L^2}^2 + \left\|\frac{1}{r}\partial_\phi\nabla_y f\right\|_{L^2}^2$$

for all $f \in H^1(\mathrm{D}\times\mathrm{D}')$.

Proof. First of all let us fix some $x \in \mathrm{D}$ and notice that, as pointed in the Remark, Lemma 5.6 is also applicable to vector-valued functions. In particular we want to apply it to the function

$$g = \begin{pmatrix} d_{\mathrm{D}}\partial_r f \\ \frac{1}{r}\partial_\phi f\end{pmatrix}$$

so that we get

$$\begin{aligned}\|d_{\mathrm{D}}\partial_r f\|_{L^2(\mathrm{D}')}^2 + \left\|\frac{1}{r}\partial_\phi f\right\|_{L^2(\mathrm{D}')}^2 &\lesssim \int_{\mathrm{D}'} d_{\mathrm{ring}}^3\left(d_{\mathrm{D}}^2|\partial_r f|^2 + \left|\frac{1}{r}\partial_\phi f\right|^2\right)dy \\ &\quad + \|d_{\mathrm{D}}\partial_r\nabla_y f\|_{L^2(\mathrm{D}')}^2 + \left\|\frac{1}{r}\partial_\phi\nabla_y f\right\|_{L^2(\mathrm{D}')}^2.\end{aligned}$$

We need to work a bit on the first addend in the right hand side, so first of all notice that of course

$$|\nabla_x f|^2 = |\partial_r f|^2 + \left|\frac{1}{r}\partial_\phi f\right|^2,$$

and second notice that

$$\nabla_y u^T \nabla_x f = \nabla_y u_r \partial_r f + \frac{1}{r}\nabla_y u_\phi \partial_\phi f.$$

Now we need to deal with the degeneracies of $\nabla_y u$, but notice that thanks to Lemma 5.7 we know that on $\partial\mathrm{D}$, the degeneracy occurs exclusively in the radial component of the vector field, namely $|\nabla_y u_r| \sim d_{\mathrm{D}}$.

On the one hand, in order to deal with the degeneracy coming from the ring we need to make use of Lemma 5.6 and use a Poincaré-Hardy argument. In particular, far from $\partial\mathrm{D}$ there holds

$$d_{\mathrm{ring}}^2 |\xi|^2 \lesssim |\nabla_y u^T \xi|^2$$

for all $\xi \in \mathbb{R}^2$, where we are using the usual estimate

$$|\nabla_y u^T \xi|^2 = \xi \cdot \nabla_y u \nabla_y u^T \xi \geq \frac{(\det \nabla_y u)^2}{\mathrm{tr}(\nabla_y u \nabla_y u^T)} |\xi|^2 \gtrsim (\det \nabla_y u)^2 |\xi|^2$$

for all $\xi \in \mathbb{R}^2$. On the other hand, in order to deal with the degeneracy at $\partial\mathrm{D}$ we have that far from the ring degeneracy and near $\partial\mathrm{D}$, there holds

$$d_{\mathrm{D}}^2 |\xi|^2 \lesssim |\nabla_y u^T \xi|^2$$

for all $\xi \in \mathbb{R}^2$. Since we are choosing $r < 1$, the ring and $\partial\mathrm{D}$ will not intersect so we can combine both arguments in the general estimate

$$d_{\mathrm{ring}}^2(x,y) d_{\mathrm{D}}^2(x) |\xi|^2 \lesssim |\nabla_y u^T(x,y)\xi|^2$$

for all $\xi \in \mathbb{R}^2$ and all $(x,y) \in \mathrm{D} \times \mathrm{D}'$.

Coming back to the estimate that we want to study, we can write

$$d_{\mathrm{D}}^2 |\partial_r f|^2 + \left|\frac{1}{r}\partial_\phi f\right|^2 \lesssim d_{\mathrm{D}}^2 |\nabla_x f|^2,$$

and thus

$$\int_{\mathrm{D}'} d_{\mathrm{ring}}^3 \left(d_{\mathrm{D}}^2 |\partial_r f|^2 + \left|\frac{1}{r}\partial_\phi f\right|^2 \right) dy \lesssim \int_{\mathrm{D}'} d_{\mathrm{ring}}^3 d_{\mathrm{D}}^2 |\nabla_x f|^2 dy \lesssim \int_{\mathrm{D}'} d_{\mathrm{ring}} |\nabla_y u^T \xi|^2 dy.$$

Finally notice that since $d_{\text{ring}} \leq d_{\text{D}'}$, the claim of the Proposition follows. ■

Without further ado we can proceed to study hypocoercivity of the one-point operator $\mathcal{L}_1 = u \cdot \nabla_x - \nu \Delta_y$ with Neumann homogeneous boundary conditions in the variable y as we did for the precious example. The strategy will be as for the shear flows example: we will define a suitable augmented energy functional $\Phi(f)$ that is coercive, namely it decays instantaneously, and that is equivalent to an appropriate H^1 norm. In this case, since we have to treat the case with boundary and since we have to deal with a degeneracy of the vector field when $x \to \partial \text{D}$, the H^1 norm considered will include a weight in the radial direction of the x gradient, namely it will be as follows,

$$\|f\|_{H^1_b}^2 = \|f\|_{L^2(\text{D}\times\text{D}')}^2 + \|\nabla_y f\|_{L^2(\text{D}\times\text{D}')} + \|d_{\text{D}} \partial_r f\|_{L^2(\text{D}\times\text{D}')}^2 + \left\|\frac{1}{r}\partial_\phi f\right\|_{L^2(\text{D}\times\text{D}')}^2 .$$

Recall that $d_{\text{D}}(x) = 1 - |x|^2 \equiv 1 - r^2$ is the distance to the boundary in $x \in \text{D}$.

For the sake of a clearer notation, we will simply denote by $\|\cdot\|$ the L^2 norm in $\text{D} \times \text{D}'$ and analogously we write $\langle \cdot, \cdot \rangle$ for the L^2 scalar product in $\text{D} \times \text{D}'$. Then, we have the following result regarding the augmented energy functional.

Lemma 5.8. *Let f be a solution to* (5.23) *and consider the augmented energy functional*

$$\begin{aligned}\Phi(f) = \frac{1}{2}\|f\|^2 + \frac{\alpha}{2}\|\nabla_y f\|^2 + \beta\langle d_{\text{D}'}\nabla_y f, \nabla_y u^T \nabla_x f\rangle \\ + \frac{\gamma}{2}\|d_{\text{D}}\partial_r f\|^2 + \frac{\gamma}{2}\left\|\frac{1}{r}\partial_\phi f\right\|^2\end{aligned} \tag{5.24}$$

with $\alpha, \beta, \gamma > 0$. Then there holds

$$\begin{aligned}\frac{d}{dt}\Phi(f) = &-\nu\|\nabla_y f\|^2 - \alpha\nu\|\Delta_y f\|^2 - \beta\|\sqrt{d_{\text{D}'}}\nabla_y u^T \nabla_x f\|^2 - \gamma\nu\|d_{\text{D}}\partial_r \nabla_y^2 f\|^2 \\ &- \gamma\nu\left\|\frac{1}{r}\partial_\phi \nabla_y f\right\|^2 - \alpha\langle \nabla_y f, \nabla_y u^T \nabla_x f\rangle - \gamma\langle d_{\text{D}}^2 \partial_r f, \partial_r u \cdot \nabla_x f\rangle \\ &- \frac{\gamma}{2}\langle u \cdot \nabla_x d_{\text{D}}^2 \partial_r f, \partial_r f\rangle - \gamma\langle \frac{1}{r}\partial_\phi f, \frac{1}{r}\partial_\phi u \cdot \nabla_x f\rangle - \frac{\gamma}{2}\langle u \cdot \nabla_x \frac{1}{r^2}\partial_\phi f, \partial_\phi f\rangle \\ &- \beta\langle d_{\text{D}'}\nabla_y f, M\nabla_x f\rangle - 2\beta\nu\langle d_{\text{D}'}\Delta_y f, \nabla_y(u\cdot\nabla_x)\cdot\nabla_y f\rangle \\ &- \beta\nu\langle d_{\text{D}'}\Delta_y f, \Delta_y u \cdot \nabla_x f\rangle - \beta\nu\langle \nabla_y d_{\text{D}'}\Delta_y f, \nabla_y u^T \nabla_x f\rangle,\end{aligned}$$

where $M = \nabla_y u^T \nabla_x u^T - (u \cdot \nabla_x)\nabla_y u^T$ is the matrix defined in Lemma 5.7.

Proof. Following a similar procedure to Lemma 5.4 we will take time derivatives in each addend from the definition of $\Phi(f)$ and use the fact that f is a solution to (5.23). First of all,

$$\frac{1}{2}\frac{d}{dt}\|f\|^2 = \langle f, \partial_t f\rangle = -\langle f, u\cdot\nabla_x f\rangle + \nu\langle f, \Delta_y f\rangle = -\nu\|\nabla_y f\|^2,$$

after integration by parts and using that $\nabla_x \cdot u = 0$. Analogously for the α term we have,

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\|\nabla_y f\|^2 &= \langle \nabla_y f, \nabla_y \partial_t f\rangle \\ &= -\langle \nabla_y f, u\cdot\nabla_x\nabla_y f\rangle - \langle \nabla_y f, \nabla_y u^T \nabla_x f\rangle + \nu\langle \nabla_y f, \Delta_y \nabla_y f\rangle \\ &= -\nu\|\Delta_y f\|^2 - \langle \nabla_y f, \nabla_y u^T\nabla_x f\rangle,\end{aligned}$$

where we use the fact that $n_y \cdot \nabla_y f = 0$ for all $x \in \mathrm{D}$ and all $y \in \mathrm{D}'$. For the γ terms we argue in a similar way,

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\|d_{\mathrm{D}}\partial_r f\|^2 &= \langle d_{\mathrm{D}}^2 \partial_r f, \partial_r \partial_t f\rangle \\ &= -\langle d_{\mathrm{D}}^2\partial_r f, u\cdot\nabla_x \partial_r f\rangle - \langle d_{\mathrm{D}}^2\partial_r f, \partial_r u\cdot\nabla_x f\rangle + \nu\langle d_{\mathrm{D}}^2\partial_r f, \Delta_y\partial_r f\rangle \\ &= -\nu\|d_{\mathrm{D}}\partial_r\nabla_y f\|^2 - \langle d_{\mathrm{D}}^2\partial_r f, \partial_r u\cdot\nabla_x f\rangle + \frac{1}{2}\langle u\cdot\nabla_x d_{\mathrm{D}}^2\partial_r f, \partial_r f\rangle,\end{aligned}$$

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\left\|\frac{1}{r}\partial_\phi f\right\|^2 &= \langle \frac{1}{r^2}\partial_\phi f, \partial_\phi\partial_t f\rangle \\ &= -\langle \frac{1}{r^2}\partial_\phi f, u\cdot\nabla_x\partial_\phi f\rangle - \langle\frac{1}{r^2}\partial_\phi f, \partial_\phi u\cdot\nabla_x f\rangle + \nu\langle\frac{1}{r^2}\partial_\phi f, \Delta_y\partial_\phi f\rangle \\ &= -\nu\left\|\frac{1}{r}\partial_r\nabla_y f\right\|^2 - \langle\frac{1}{r^2}\partial_\phi f, \partial_\phi u\cdot\nabla_x f\rangle + \frac{1}{2}\langle u\cdot\nabla_x\frac{1}{r^2}\partial_\phi f, \partial_\phi f\rangle.\end{aligned}$$

Finally for the β terms we have the following,

$$\begin{aligned}\frac{d}{dt}\langle d_{\mathrm{D}'}\nabla_y f, \nabla_y u^T\nabla_x f\rangle &= \langle d_{\mathrm{D}'}\nabla_y\partial_t f, \nabla_y u^T\nabla_x f\rangle + \langle d_{\mathrm{D}'}\nabla_y f, \nabla_y u^T\nabla_x\partial_t f\rangle \\ &= -\langle d_{\mathrm{D}'}\nabla_y(u\cdot\nabla_x f), \nabla_y u^T\nabla_x f\rangle - \langle d_{\mathrm{D}'}\nabla_y f, \nabla_y u^T\nabla_x(u\cdot\nabla_x f)\rangle \\ &\quad + \nu\langle d_{\mathrm{D}'}\nabla_y\Delta_y f, \nabla_y u^T\nabla_x f\rangle + \nu\langle d_{\mathrm{D}'}\nabla_y f, \nabla_y u^T\nabla_x\Delta_y f\rangle \\ &= \mathrm{I} + \nu\,\mathrm{II}\,.\end{aligned}$$

On the one hand the terms without ν can be rearranged as

$$\begin{aligned}
\mathrm{I} &= -\|\sqrt{d_{\mathrm{D}'}}\nabla_y u^T \nabla_x f\|^2 - \langle d_{\mathrm{D}'}(u\cdot\nabla_x)\nabla_y f, \nabla_y u^T \nabla_x f\rangle \\
&\quad - \langle d_{\mathrm{D}'}\nabla_y f, \nabla_y u^T \nabla_x u^T \nabla_x f\rangle - \sum_i \langle d_{\mathrm{D}'}\nabla_y f, u_i \nabla_y u^T \nabla_x \partial_{x_i} f\rangle \\
&= -\|\sqrt{d_{\mathrm{D}'}}\nabla_y u^T \nabla_x f\|^2 + \langle d_{\mathrm{D}'}\nabla_y f, (u\cdot\nabla_x)\nabla_y u^T \nabla_x f\rangle \\
&\quad + \sum_i \langle d_{\mathrm{D}'}\nabla_y f, u_i \nabla_y u^T \nabla_x \partial_{x_i} f\rangle - \langle d_{\mathrm{D}'}\nabla_y f, \nabla_y u^T \nabla_x u^T \nabla_x f\rangle \\
&\quad - \sum_i \langle d_{\mathrm{D}'}\nabla_y f, u_i \nabla_y u^T \nabla_x \partial_{x_i} f\rangle \\
&= -\|\sqrt{d_{\mathrm{D}'}}\nabla_y u^T \nabla_x f\|^2 - \langle d_{\mathrm{D}'}\nabla_y f, \nabla_y u^T \nabla_x u^T \nabla_x f\rangle \\
&\quad + \langle d_{\mathrm{D}'}\nabla_y f, (u\cdot\nabla_x)\nabla_y u^T \nabla_x f\rangle \\
&= -\|\sqrt{d_{\mathrm{D}'}}\nabla_y u^T \nabla_x f\|^2 - \langle d_{\mathrm{D}'}\nabla_y f, M\nabla_x f\rangle,
\end{aligned}$$

where we define $M = \nabla_y u^T \nabla_x u^T - (u\cdot\nabla_x)\nabla_y u^T$ as in Lemma 5.7. On the other hand, for the addends with ν we have,

$$\begin{aligned}
\mathrm{II} &= -\langle d_{\mathrm{D}'}\Delta_y f, \Delta_y u\cdot\nabla_x f\rangle - \langle d_{\mathrm{D}'}\Delta_y f, \nabla_y(u\cdot\nabla_x)\cdot\nabla_y f\rangle \\
&\quad - \langle \nabla_y d_{\mathrm{D}'}\Delta_y f, \nabla_y u^T \nabla_x f\rangle - \langle d_{\mathrm{D}'}\nabla_y(u\cdot\nabla_x)\cdot\nabla_y f, \Delta_y f\rangle,
\end{aligned}$$

where notice that we can perform the integration by parts with ∇_y only because the term $d_{\mathrm{D}'}$ vanishes at D'. All in all, putting all the terms together we arrive to the claim of the Lemma. ■

Lemma 5.9. *Let f be a solution to* (5.23) *and $\Phi(f)$ be the augmented energy functional from* (5.24)*. There exists a choice of $\alpha, \beta, \gamma > 0$ such that:*

1. *$\Phi(\cdot)$ is equivalent to $\|\cdot\|^2_{H^1_b}$ as a norm in $\mathrm{D}\times\mathrm{D}'$, namely*

$$\min\left\{1, \frac{1}{\nu^2}\right\}\|f\|^2_{H^1_b} \lesssim \Phi(f) \lesssim \max\left\{1, \frac{1}{\nu^2}\right\}\|f\|^2_{H^1_b}.$$

2. *If $\nu > 1$ is sufficiently large, then there exist $\delta_1, \delta_2 > 0$ independent of ν so that there holds*

$$\begin{aligned}
\frac{d}{dt}\Phi(f) + \frac{\nu}{2}\|\nabla_y f\|^2 + \frac{\delta_1}{\nu}\|\sqrt{d_{\mathrm{D}'}}\nabla_y u^T \nabla_x f\|^2 \\
+ \frac{\delta_2}{\nu}\|d_{\mathrm{D}}\partial_r \nabla_y f\|^2 + \frac{\delta_2}{\nu}\left\|\frac{1}{r}\partial_\phi \nabla_y f\right\|^2 \le 0
\end{aligned}$$

for all $t > 0$.

Proof. In a similar fashion to the case of the shear flow, we can prove first of all the equivalence between Φ and $\|\cdot\|^2_{H^1_b}$ via Hölder and Young's inequalities. Indeed, recalling the definition of the augmented energy functional,

$$\Phi(f) = \frac{1}{2}\|f\|^2 + \frac{\alpha}{2}\|\nabla_y f\|^2 + \beta\langle d_{\mathrm{D}'}\nabla_y f, \nabla_y u^T \nabla_x f\rangle + \frac{\gamma}{2}\left(\|d_{\mathrm{D}}\partial_r f\|^2 + \left\|\frac{1}{r}\partial_\phi f\right\|^2\right),$$

then, taking into account that $d_{\mathrm{D}'} \leq r$, we can write

$$\beta|\langle d_{\mathrm{D}'}\nabla_y f, \nabla_y u^T \nabla_x f\rangle| \leq \frac{\beta r}{2}\sqrt{\frac{\alpha}{\gamma}}\|\nabla_y f\|^2 + \frac{\beta r}{2}\sqrt{\frac{\gamma}{\alpha}}\|\nabla_y u^T \nabla_x f\|^2.$$

But using the properties of the vector field from Lemma 5.7 we obtain that

$$\|\nabla_y u^T \nabla_x f\|^2 = \|\nabla_y u_r \partial_r f\|^2 + \left\|\frac{1}{r}\nabla_y u_\phi \partial_\phi f\right\|^2 \leq \|d_{\mathrm{D}}\partial_r f\|^2 + \left\|\frac{1}{r}\partial_\phi f\right\|^2,$$

and thus

$$\beta|\langle d_{\mathrm{D}'}\nabla_y f, \nabla_y u^T \nabla_x f\rangle| \leq \frac{\beta r}{2}\sqrt{\frac{\alpha}{\gamma}}\|\nabla_y f\|^2 + \frac{\beta r}{2}\sqrt{\frac{\gamma}{\alpha}}\left(\|d_{\mathrm{D}}\partial_r f\|^2 + \left\|\frac{1}{r}\partial_\phi f\right\|^2\right).$$

In this way we obtain that the augmented energy functional is bounded by,

$$\Phi(f) \geq \frac{1}{2}\|f\|^2 + \frac{\alpha}{2}\left(1 - \frac{\beta r}{\sqrt{\alpha\gamma}}\right)\|\nabla_y f\|^2 + \frac{\gamma}{2}\left(1 - \frac{\beta r}{\sqrt{\alpha\gamma}}\right)\left(\|d_{\mathrm{D}}\partial_r f\|^2 + \left\|\frac{1}{r}\partial_\phi f\right\|^2\right),$$

$$\Phi(f) \leq \frac{1}{2}\|f\|^2 + \frac{\alpha}{2}\left(1 + \frac{\beta r}{\sqrt{\alpha\gamma}}\right)\|\nabla_y f\|^2 + \frac{\gamma}{2}\left(1 + \frac{\beta r}{\sqrt{\alpha\gamma}}\right)\left(\|d_{\mathrm{D}}\partial_r f\|^2 + \left\|\frac{1}{r}\partial_\phi f\right\|^2\right),$$

and therefore $\Phi(f)$ and $\|\cdot\|^2_{H^1_b}$ are equivalent provided that $\beta r < \sqrt{\alpha\gamma}$.

Notice that in order to obtain the condition $\beta r < \sqrt{\alpha\gamma}$, we have assumed that $\|\nabla_y u\|_\infty \leq 1$ and $\|\nabla_y u_\phi\|_\infty \leq 1$. In principle, given the explicit definition of our vector field in (5.22), we should keep track of these constants or define some $K > 0$ such that the L^∞ norm of all components and derivatives of u is bounded by it. Another option could be to smuggle $K > 0$ in in definition of u such that the correct bound holds directly by definition. We will not do this for the sake of a clearer notation and because nothing changes in the end, the argument would be exactly the same only keeping track of K.

The second point to study for this Lemma concerns the stability estimate in *2*. We must find a control over all the terms that do not have a sign in Lemma 5.8 with the dissipation terms, namely the terms that do have a (correct) sign. We use a combination of Hölder and Young's inequalities repeatedly in all these terms. In order to bound correctly many of the terms we need to make use of the Poincaré-Hardy estimate from Proposition 5.2. We start with the α term.

$$\alpha|\langle \nabla_y f, \nabla_y u^T \nabla_x f\rangle| \leq \frac{\nu}{2}\|\nabla_y f\|^2 + \frac{\alpha^2}{2\nu}\|\nabla_y u^T \nabla_x f\|^2.$$

Then, using the properties $\|\nabla_y u_r\|_\infty \leq d_{\mathrm{D}}$ and $\|\nabla_y u_\phi\|_\infty \leq 1$ we get

$$\begin{aligned}
\alpha|\langle \nabla_y f, \nabla_y u^T \nabla_x f\rangle| &\leq \frac{\nu}{2}\|\nabla_y f\|^2 + \frac{\alpha^2}{2\nu}\|\nabla_y u_r \partial_r f\|^2 + \frac{\alpha^2}{2\nu}\left\|\frac{1}{r}\nabla_y u_\phi \partial_\phi f\right\|^2 \\
&\leq \frac{\nu}{2}\|\nabla_y f\|^2 + \frac{\alpha^2}{2\nu}\left(\|d_{\mathrm{D}}\partial_r f\|^2 + \left\|\frac{1}{r}\partial_\phi f\right\|^2\right) \\
&\leq \frac{\nu}{2}\|\nabla_y f\|^2 \\
&\quad + \frac{\alpha^2 c_{PH}}{2\nu}\left(\|\sqrt{d_{\mathrm{D}'}}\nabla_y u^T \nabla_x f\|^2 + \|d_{\mathrm{D}}\partial_r \nabla_y f\|^2 + \left\|\frac{1}{r}\partial_\phi \nabla_y f\right\|^2\right).
\end{aligned}$$

In the last step we have used Proposition 5.2 with a Poincaré-Hardy constant $c_{PH} > 0$ instead of $\lesssim$ so that we can carefully keep track of all the terms involved. Analogously for the γ terms we have,

$$\begin{aligned}
\gamma|\langle d_{\mathrm{D}}^2 \partial_r f, \partial_r u \cdot \nabla_x f\rangle| &\leq \gamma\|d_{\mathrm{D}}\partial_r f\|\left(\|d_{\mathrm{D}}\partial_r f\| + \left\|\frac{1}{r}\partial_\phi f\right\|\right) \\
&\leq \frac{3\gamma}{2}\|d_{\mathrm{D}}\partial_r f\|^2 + \frac{\gamma}{2}\left\|\frac{1}{r}\partial_\phi f\right\|^2 \\
&\leq \frac{3\gamma c_{PH}}{2}\left(\|\sqrt{d_{\mathrm{D}'}}\nabla_y u^T \nabla_x f\|^2 + \|d_{\mathrm{D}}\partial_r \nabla_y f\|^2 + \left\|\frac{1}{r}\partial_\phi \nabla_y f\right\|^2\right).
\end{aligned}$$

For the next one we must take into account that $u \cdot \nabla_x d_{\mathrm{D}} = -4 d_{\mathrm{D}} u \cdot x = -4 d_{\mathrm{D}} |x| u_r$. Therefore,

$$\begin{aligned}
\frac{\gamma}{2}|\langle u \cdot \nabla_x d_{\mathrm{D}}^2 \partial_r f, \partial_r f\rangle| &= 2\gamma|\langle d_{\mathrm{D}}|x|u_r \partial_r f, \partial_r f\rangle| \leq 2\gamma\|d_{\mathrm{D}}\partial_r f\|^2 \\
&\leq 2\gamma c_{PH}\left(\|\sqrt{d_{\mathrm{D}'}}\nabla_y u^T \nabla_x f\|^2 + \|d_{\mathrm{D}}\partial_r \nabla_y f\|^2 + \left\|\frac{1}{r}\partial_\phi \nabla_y f\right\|^2\right),
\end{aligned}$$

where we have used again $\|u_r\|_\infty \leq d_{\mathrm{D}}$.

$$\begin{aligned}
\gamma\left|\langle\frac{1}{r}\partial_\phi f, \frac{1}{r}\partial_\phi u\cdot\nabla_x f\rangle\right| &\leq \gamma\left\|\frac{1}{r}\partial_\phi f\right\|\left(\left\|\frac{1}{r}\partial_\phi u_r\partial_r f\right\| + \left\|\frac{1}{r^2}\partial_\phi u_\phi\partial_\phi f\right\|\right)\\
&\leq \gamma\left\|\frac{1}{r}\partial_\phi f\right\|\left(\|d_{\mathrm{D}}\partial_r f\| + \left\|\frac{1}{r}\partial_\phi f\right\|\right)\\
&\leq \frac{\gamma}{2}\|d_{\mathrm{D}}\partial_r f\|^2 + \frac{3\gamma}{2}\left\|\frac{1}{r}\partial_\phi f\right\|^2\\
&\leq \frac{3\gamma c_{PH}}{2}\left(\|\sqrt{d_{\mathrm{D}'}}\nabla_y u^T\nabla_x f\|^2 + \|d_{\mathrm{D}}\partial_r\nabla_y f\|^2 + \left\|\frac{1}{r}\partial_\phi\nabla_y f\right\|^2\right),
\end{aligned}$$

where yet again we use the properties of the vector field. The last term with γ is controlled analogously and taking into account that

$$u\cdot\nabla_x\frac{1}{r^2} = -\frac{2}{r^3}u_r \quad \text{and} \quad \left\|\frac{u_r}{r}\right\|_\infty \leq 1.$$

Hence,

$$\begin{aligned}
\frac{\gamma}{2}\left|\langle u\cdot\nabla_x\frac{1}{r^2}\partial_\phi f, \partial_\phi f\rangle\right| &\leq \gamma\left\|\frac{1}{r}\partial_\phi f\right\|^2\\
&\leq \gamma c_{PH}\left(\|\sqrt{d_{\mathrm{D}'}}\nabla_y u^T\nabla_x f\|^2 + \|d_{\mathrm{D}}\partial_r\nabla_y f\|^2 + \left\|\frac{1}{r}\partial_\phi\nabla_y f\right\|^2\right).
\end{aligned}$$

Finally, for the β terms we follow the same philosophy. To begin with, we use yet again Lemma 5.7 and the properties of the matrix M,

$$\begin{aligned}
\beta|\langle d_{\mathrm{D}'}\nabla_y f, M\nabla_x f\rangle| &\leq \frac{\nu}{4}\|\nabla_y f\|^2 + \frac{\beta^2 r^2}{\nu}\|M\nabla_x f\|^2\\
&\leq \frac{\nu}{4}\|\nabla_y f\|^2 + \frac{\beta^2 r^2 c_M}{\nu}\left(\|d_{\mathrm{D}}\partial_r f\|^2 + \left\|\frac{1}{r}\partial_\phi f\right\|^2\right)\\
&\leq \frac{\nu}{4}\|\nabla_y f\|^2 + \frac{\beta^2 r^2 c_M c_{PH}}{\nu}\|\sqrt{d_{\mathrm{D}'}}\nabla_y u^T\nabla_x f\|^2\\
&\quad + \frac{\beta^2 r^2 c_M c_{PH}}{\nu}\left(\|d_{\mathrm{D}}\partial_r\nabla_y f\|^2 + \left\|\frac{1}{r}\partial_\phi\nabla_y f\right\|^2\right),
\end{aligned}$$

where again, in order to be precise with all the constants, we introduce $c_M > 0$ from

Lemma 5.7 instead of writing $\lesssim$. Next in order,

$$\begin{aligned}2\beta\nu|\langle d_{\mathrm{D}'}\Delta_y f, \nabla_y(u\cdot\nabla_x)\cdot\nabla_y f\rangle| &\leq \frac{\alpha\nu}{4}\|\Delta_y f\|^2 + \frac{4\beta^2\nu r^2}{\alpha}\|\nabla_y(u\cdot\nabla_x)\cdot\nabla_y f\|^2 \\ &\leq \frac{\alpha\nu}{4}\|\Delta_y f\|^2 + \frac{4\beta^2\nu r^2}{\alpha}\|\nabla_y u_r\cdot\partial_r\nabla_y f\|^2 \\ &\quad + \frac{4\beta^2\nu r^2}{\alpha}\left\|\frac{1}{r}\nabla_y u_\phi\cdot\partial_\phi\nabla_y f\right\|^2 \\ &\leq \frac{\alpha\nu}{4}\|\Delta_y f\|^2 + \frac{4\beta^2\nu r^2}{\alpha}\left(\|d_{\mathrm{D}}\partial_r\nabla_y f\|^2 + \left\|\frac{1}{r}\partial_\phi\nabla_y f\right\|^2\right)\end{aligned}$$

where again we used that u and all its derivatives are bounded by 1. For the next term we need to use the properties of u as in Lemma 5.7, namely

$$\|\Delta_y u\cdot\nabla_x f\|^2 = \|\Delta_y u_r\partial_r f\|^2 + \left\|\frac{1}{r}\Delta_y u_\phi\partial_\phi f\right\|^2 \leq \|d_{\mathrm{D}}\partial_r f\|^2 + \left\|\frac{1}{r}\partial_\phi f\right\|^2.$$

Therefore a direct application of Poincaré-Hardy from Proposition 5.2 gives,

$$\|\Delta_y u\cdot\nabla_x f\|^2 \leq c_{PH}\left(\|\sqrt{d_{\mathrm{D}'}}\nabla_y u^T\nabla_x f\|^2 + \|d_{\mathrm{D}}\partial_r\nabla_y f\|^2 + \left\|\frac{1}{r}\partial_\phi\nabla_y f\right\|^2\right),$$

so that the next term is controlled by

$$\begin{aligned}\beta\nu|\langle d_{\mathrm{D}'}\Delta_y f, \Delta_y u\cdot\nabla_x f\rangle| &\leq \frac{\alpha\nu}{4}\|\Delta_y f\|^2 + \frac{\beta^2\nu r^2}{\alpha}\|\Delta_y u\cdot\nabla_x f\|^2 \\ &\leq \frac{\alpha\nu}{4}\|\Delta_y f\|^2 + \frac{\beta^2\nu r^2 c_{PH}}{\alpha}\|\sqrt{d_{\mathrm{D}'}}\nabla_y u^T\nabla_x f\|^2 \\ &\quad + \frac{\beta^2\nu r^2 c_{PH}}{\alpha}\left(\|d_{\mathrm{D}}\partial_r\nabla_y f\|^2 + \left\|\frac{1}{r}\partial_\phi\nabla_y f\right\|^2\right).\end{aligned}$$

The last term is concerned with the y gradient of $d_{\mathrm{D}'}$, but this can explicitly computed

$$\nabla_y d_{\mathrm{D}'} = \nabla_y\left(\frac{r^2-|y|^2}{r}\right) = -2\frac{y}{r},$$

and controlled by $\|\nabla_y d_{\mathrm{D}'}\|_\infty \leq 2$. Thus, by the same argument as the previous terms we get

$$\begin{aligned}\beta\nu|\langle \nabla_y d_{\mathrm{D}'}\Delta_y f, \nabla_y u^T\nabla_x f\rangle| &\leq \frac{\alpha\nu}{4}\|\Delta_y f\|^2 + \frac{4\beta^2\nu}{\alpha}\|\nabla_y u^T\nabla_x f\|^2\\ &\leq \frac{\alpha\nu}{4}\|\Delta_y f\|^2 + \frac{4\beta^2\nu c_{PH}}{\alpha}\|\sqrt{d_{\mathrm{D}'}}\nabla_y u^T\nabla_x f\|^2\\ &\quad + \frac{4\beta^2\nu c_{PH}}{\alpha}\left(\|d_{\mathrm{D}}\partial_r\nabla_y f\|^2 + \left\|\frac{1}{r}\partial_\phi\nabla_y f\right\|^2\right).\end{aligned}$$

All in all, using these estimates in the expression from Lemma 5.8 we get

$$\begin{aligned}\frac{d}{dt}\Phi(f) + \frac{\nu}{4}\|\nabla_y f\|^2 + \frac{\alpha\nu}{4}\|\Delta_y f\|^2 + \mathrm{I}(\nu)\|\sqrt{d_{\mathrm{D}'}}\nabla_y u^T\nabla_x f\|^2&\\ + \mathrm{II}(\nu)\|d_{\mathrm{D}}\partial_r\nabla_y f\|^2 + \mathrm{III}(\nu)\left\|\frac{1}{r}\partial_\phi\nabla_y f\right\|^2 &\leq 0,\end{aligned}$$

with coefficients

$$\mathrm{I}(\nu) = \beta - \frac{\alpha^2 c_{PH}}{2\nu} - 6\gamma c_{PH} - \frac{\beta^2 r^2 c_M c_{PH}}{\nu} - \frac{\beta^2\nu(r^2+4)c_{PH}}{\alpha},$$

$$\mathrm{II}(\nu) = \mathrm{III}(\nu) = \gamma\nu - \frac{\alpha^2 c_{PH}}{2\nu} - 6\gamma c_{PH} - \frac{\beta^2 r^2 c_M c_{PH}}{\nu} - \frac{\beta^2\nu(5r^2+4)c_{PH}}{\alpha}.$$

In order to obtain the desired result, we need to choose $\alpha, \beta, \gamma > 0$ such that $\mathrm{I}(\nu), \mathrm{II}(\nu), \mathrm{II}(\nu) > 0$ and with $\beta r < \sqrt{\alpha\gamma}$. So let us make the following ansatz,

$$\alpha = a, \quad \beta = \frac{b}{\nu}, \quad \gamma = \frac{c}{\nu^2}.$$

Then we obtain the following compatibility conditions.

1. The equivalence between Φ and $\|\cdot\|^2_{H^1_b}$ can be achieved by assuming

$$b^2 < ac. \tag{5.25}$$

2. Positivity of $\mathrm{I}(\nu)$ can be achieved by assuming

$$\frac{b}{\nu} > \frac{c_{PH}}{\nu}\left(\frac{a^2}{2} + \frac{5b^2}{a}\right) + \frac{1}{\nu^2}\left(6cc_{PH} + b^2 r^2 c_M c_{PH}\right),$$

and therefore this holds if

$$b > c_{PH}\left(\frac{a^2}{2} + \frac{5b^2}{a}\right) \tag{5.26}$$

and if $\nu > 0$ is big enough.

3. Positivity of II(ν) and III(ν) can be achieved by assuming

$$\frac{c}{\nu} > \frac{c_{PH}}{\nu}\left(\frac{a^2}{2} + \frac{9b^2}{a}\right) + \frac{1}{\nu^2}\left(6cc_{PH} + b^2r^2c_Mc_{PH}\right),$$

and therefore this holds if

$$c > c_{PH}\left(\frac{a^2}{2} + \frac{9b^2}{a}\right) \tag{5.27}$$

and if $\nu > 0$ is big enough.

Notice that in general condition (5.27) implies (5.25), at least provided that $c_{PH} > 1/9$, which can always be assumed. Therefore we obtain a system that is solvable. In particular we can choose the constants $a, b, c > 0$ such that

$$0 < a < \frac{1}{10c_{PH}^2},$$

$$\frac{a}{5c_{PH}}\left(1 - \sqrt{1 - 10ac_{PH}^2}\right) < b < \frac{a}{5c_{PH}}\left(1 + \sqrt{1 - 10ac_{PH}^2}\right),$$

$$c > c_{PH}\left(\frac{a^2}{2} + \frac{9b^2}{a}\right).$$

With these constants $a, b, c > 0$ and $\nu > 0$ we finally obtain the desired claim of the Lemma. ■

As in the previous example, we have now in our hands all the needed tools to prove hypocoercivity of the operator $\mathcal{L}_1 = -u \cdot \nabla_x + \nu\Delta_y$ and homogeneous Neumann boundary conditions in $\mathrm{D} \times \mathrm{D}'$ with this particular vector field u.

Proposition 5.3. *Let $f(t, \cdot)$ be a solution to* (5.23) *in* $\mathrm{D} \times \mathrm{D}'$ *and assume that $\nu > 1$ is sufficiently large. Then there holds*

$$\|f(t, \cdot)\|_{H_b^1} \lesssim \nu \|f(0, \cdot)\|_{H_b^1} e^{-\frac{1}{2\nu}t}$$

for all $t > 0$ and where the constants absorbed in $\lesssim$ do not depend on ν.

Proof. We start using the decay of the augmented energy functional from Lemma 5.9.

$$\frac{d}{dt}\Phi(f) + \frac{\nu}{2}\|\nabla_y f\|^2 + \frac{\delta_1}{\nu}\|\sqrt{d_{\mathrm{D}'}}\nabla_y u^T \nabla_x f\|^2 + \frac{\delta_2}{\nu}\left(\|d_{\mathrm{D}}\partial_r \nabla_y f\|^2 + \left\|\frac{1}{r}\partial_\phi \nabla_y f\right\|^2\right) \leq 0,$$

and together with the Poincaré-Hardy inequality from Proposition 5.2 there yields,

$$\frac{d}{dt}\Phi(f) + \frac{\nu}{2}\|\nabla_y f\|^2 + \frac{1}{\nu}\left(\|d_{\mathrm{D}}\partial_r f\|^2 + \left\|\frac{1}{r}\partial_\phi f\right\|^2\right) \lesssim 0.$$

We need to produce a factor $\|f\|^2$ in that expression in order to use the equivalence between Φ and $\|\cdot\|^2_{H^1_b}$, but now we cannot apply a standard Poincaré inequality because the norms have a weight. In order to deal with this issue we need to make use of a suitable *weighted Poincaré inequality* of the form

$$\|f\|^2 \lesssim \|\nabla_y f\|^2 + \|d_{\mathrm{D}}\nabla_x f\|^2 \leq \|\nabla_y f\|^2 + \|d_{\mathrm{D}}\partial_r f\|^2 + \left\|\frac{1}{r}\partial_\phi f\right\|^2,$$

where we are taking into account that f has mean zero for all $t > 0$. See Section A.2 on the Appendix for more details about this weighted Poincaré inequality. Then, we obtain

$$\frac{d}{dt}\Phi(f) + \frac{1}{\nu}\|f\|^2 + \nu\|\nabla_y f\|^2 + \frac{1}{\nu}\left(\|d_{\mathrm{D}}\partial_r f\|^2 + \left\|\frac{1}{r}\partial_\phi f\right\|^2\right) \lesssim 0.$$

and thus

$$\frac{d}{dt}\Phi(f) + \frac{1}{\nu}\|f\|^2_{H^1_b} \lesssim 0.$$

Now we can use the equivalence between Φ and $\|\cdot\|^2_{H^1_b}$ from Lemma 5.9, and since we are assuming that $\nu > 1$ is big enough, the equivalence result reads

$$\frac{1}{\nu^2}\|f\|^2_{H^1_b} \lesssim \Phi(f) \lesssim \|f\|^2_{H^1_b}.$$

Namely we obtain that

$$\frac{d}{dt}\Phi(f) + \frac{1}{\nu}\Phi(f) \lesssim 0,$$

and via Gronwall and yet again the equivalence from Lemma 5.9,

$$\|f(t)\|^2_{H^1_b} \lesssim \nu^2\|f(0)\|^2_{H^1_b} e^{-\frac{1}{\nu}t}$$

so that there yields the claim of the Proposition. ■

Remark 5.9. Notice that in this example we are dealing with a vortex that moves in a subset of the domain $\mathrm{D} = B_1(0)$. This subset is set to be a ball of radius $r > 0$, but we have proved that the ergodicity result holds true for all $r > 0$. Obviously we strongly use the fact that y moves in a open domain, it cannot be just one point, but what it is remarkable from here is that if $y = 0$ then only a rigid rotation occurs. However, if we let y move in a ball of radius $r \ll 1$, something that would be very close to the rigid rotation, what we obtain thanks to Theorem 5.1 is that $\|\mathbb{E}\theta\|_{L^2}$ decays exponentially fast to 0.

This is not yet a mixing result in the standard sense (pointwise in the noise) but it is a first step, and it is surprising from a physical perspective since we show that the decay will happen exponentially fast for *any* $r > 0$.

A. Additional results

A.1. Stochastic Lagrangian flows on domains with boundary

Many of the results considered along this monograph regarding the transport and advection-diffusion equations deal with the so called *Lagrangian representation of the solution* which consists of a perspective that follows the particle trajectories. Recall that for the transport equation, namely if $\kappa = 0$, we have that solutions θ to the PDE,

$$\begin{cases} \partial_t \theta + u \cdot \nabla \theta &= 0 \quad \text{in } (0,T) \times \mathrm{D}, \\ \theta(0,\cdot) &= \theta_0 \quad \text{in } \mathrm{D}, \end{cases} \tag{A.1}$$

and the particle trajectories, that are solutions X_t to the ODE,

$$\frac{dX_t}{dt} = u(t, X_t)dt, \quad X_0 = \mathrm{id}$$

are related through the method of characteristics: $\theta(t,\cdot) = (X_t)_\# \theta_0$.

If we let $\kappa > 0$, then the particle trajectories stop being a deterministic object due to the effect of the diffusion. Let θ^κ be a solution to

$$\begin{cases} \partial_t \theta^\kappa + u \cdot \nabla \theta^\kappa &= \kappa \Delta \theta^\kappa \quad \text{in } (0,T) \times \mathrm{D}, \\ \theta^\kappa(0,\cdot) &= \theta_0^\kappa \quad \text{in } \mathrm{D}, \end{cases} \tag{A.2}$$

with u smooth. If in addition we let D be a bounded domain with (smooth) boundary, the Lagrangian representation of the particles trajectories is slightly more involved. Due to the presence of the Laplacian, the Lagrangian representation will be stochastic.

Hence, we fix a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ on which we define the according stochastic process. For a domain D having for each $x \in \partial \mathrm{D}$ a unique normal $n(x)$, it is well-know (see e.g. [76, Chapter 3.1]) that any smooth solution of (A.2) is intimately related to the solution to the SDE

$$dX_t = u(t, X_t)\, ds + \sqrt{2\kappa}\, dB_t - n(X_t)\, dL_t, \tag{A.3}$$

where $\{B_t\}_{t \geq 0}$ is an $\mathcal{F}_t$-adapted Brownian motion on $\mathbb{R}^d$ and $\{L_t\}_{t \geq 0}$ is an $\mathcal{F}_t$-adapted local time of the process $\{X_t\}_{t\geq 0}$ at the boundary $\partial\mathrm{D}$, that is a non-decreasing process with $L_0 = 0$ such that

$$\int_0^t dL_s \leq t, \qquad \int_0^t \mathbb{1}_{X_s \notin \partial\mathrm{D}}\, dL_s = 0. \tag{A.4}$$

The representation is obtained via the Kolmogorov backward equation associated to (A.2), that is a solution $f : [0,t] \times \mathrm{D} \to \mathbb{R}$ of the backward parabolic equation with some terminal condition $g \in C^2(\mathrm{D})$ satisfying

$$\begin{aligned} \partial_t f + \kappa\Delta f + u \cdot \nabla f &= 0 \qquad && \text{in } [0,t] \times \mathrm{D}, \\ \nabla f(s,x) \cdot n(x) &= 0 && \text{for } (s,x) \in [0,t] \times \partial\mathrm{D}, \\ f(t,\cdot) &= g && \text{in D.} \end{aligned} \tag{A.5}$$

Then, any solution of (A.3) provides a solution to (A.5) via the observable representation

$$f(s,x) = \mathbb{E}_{s,x}[g(X(t))] = \mathbb{E}[g(X(t))|X(s) = x] \qquad \text{for } (s,x) \in (0,t) \times \mathrm{D}. \tag{A.6}$$

From here we arrive at measure-valued solutions to (A.2) via duality, which we give in the following definition.

Definition A.1 (Measure-valued solution to (A.2))**.** A Borel curve $\theta^\kappa = (\theta^\kappa_t)_{t\in[0,T]} \subset \mathcal{M}(\mathbb{R}^d)$ is a measure-valued solution to the advection-diffusion equation (A.2) provided that

$$\int_0^T \int_D (\kappa + |u(t,\cdot)|)\, d|\theta^\kappa_t(\cdot)|\, dt < \infty \tag{A.7}$$

and for all $f \in C^{1,2}([0,T] \times \overline{\mathrm{D}}) \cap \{\partial_n f \equiv 0 \text{ on } \partial\mathrm{D}\}$ and all $0 \leq t_1 \leq t_2 \leq T$ it holds

$$\int_\mathrm{D} f(t_2,\cdot)\, d\theta^\kappa_{t_2} - \int_\mathrm{D} f(t_1,\cdot)\, d\theta^\kappa_{t_1} = \int_{t_1}^{t_2} \int_\mathrm{D} (\partial_t + \kappa\Delta + u \cdot \nabla) f(t,x)\, d\theta^\kappa_t(x)\, dt = 0. \tag{A.8}$$

By a standard density argument [4, Lemma 8.1.2], there holds that any measure-valued solution in the sense of Definition A.1 admits a narrowly continuous representative, coinciding with $(\theta^\kappa_t)_{t\in(0,T)}$ for a.e. $t \in (0,T)$, in the space $\mathcal{M}(\mathbb{R}^d)$ conserving the mass, i.e. $\theta^\kappa_t(\overline{\mathrm{D}}) = \theta^\kappa_0(\overline{\mathrm{D}})$ for all $t \in (0,T]$. Hence, we can without loss of generality consider narrowly continuous paths $(\theta^\kappa_t)_{t\in(0,T)} \subset \mathcal{P}(\overline{\mathrm{D}})$ solution to the advection-diffusion equation in the sense of Definition A.1.

For smooth u, we find a unique classical solutions $f \in C^{1,2}([0,T]\times\mathrm{D})$ to the system (A.5) (see [47]) with terminal value $g \in C^2(\mathrm{D})$. In particular this identifies via (A.8), becoming for $t_1 = 0$ and $t \in (0,T]$ the identity

$$\int_{\mathrm{D}} g(\cdot)\, d\theta_t^\kappa = \int_{\mathrm{D}} f(0,\cdot)\, d\theta_0^\kappa$$

a unique family $(\theta_t^\kappa)_{t\in[0,T]} \subseteq \mathcal{P}(\mathrm{D})$.

Based on the stochastic representation (A.6), we obtain the pathwise Lagrangian representation

$$\int_{\mathrm{D}} g(\cdot)\, d\theta_t^\kappa = \int_{\mathrm{D}} \mathbb{E}_{0,x}[g(X_t)]\, d\theta_0^\kappa = \mathbb{E}[g(X_t) \,|\, \text{law } X_0 = \theta_0^\kappa] = \mathbb{E}_{\theta_0^\kappa}[g(X_t)]. \tag{A.9}$$

Remark A.1 (Stochastic Lagrangian flows). For the case of $\mathrm{D} = \mathbb{R}^d$, i.e. no reflection, the Lagrangian representation (A.9) was obtained in [45], for bounded coefficients and in [92] under the sole integrability condition (A.7). The identification, also called stochastic Lagrangian flow, is based on martingale solutions to (A.3), which is a weak solution concept for SDEs going back to [88, 89].

It seems also possible to directly generalize the concept of stochastic Lagrangian flows on a bounded set with reflecting boundary conditions to measurable vectorfields just satisfying (A.7). Here, one would use the martingale problem formulation from [90] for the reflected SDE (A.3) (see also [76, Chapter 3.2]) and do similar approximation steps as outlined in [92, Appendix A].

A.2. Weighted Poincaré inequalities

When dealing with coercive and hypocoercive properties of operators as we did in Sections 5.5 and 5.6 it is of the utmost importance to have suitable Poincaré inequalities.

One simple example of such situation would be to consider the heat equation. Let $\mathrm{D} \subset \mathbb{R}^d$ be a nice bounded domain and f be a solution to

$$\partial_t f = \kappa \Delta f \quad \text{in } (0,\infty) \times \mathrm{D}$$

with homogeneous Neumann boundary condition, $n \cdot \nabla f = 0$ in $\partial\mathrm{D}$. For this equation we have the standard energy estimate

$$\frac{1}{2}\frac{d}{dt}\|f\|_{L^2}^2 + \kappa\|\nabla f\|_{L^2}^2 = 0,$$

and if we want to translate this into a exponential decay of the L^2 norm, we need to make use of a Poincaré inequality. In particular, the standard Poincaré inequality for bounded domains with smooth boundary reads as follows

$$\left\| f - \fint f dx \right\|_{L^2} \lesssim \|\nabla f\|_{L^2}, \tag{A.10}$$

that in general can be stated with any L^p norm, see [41] for details.

For the heat equation we also know that if we consider a mean zero initial datum

$$\int_{\mathrm{D}} f(0,x)dx = 0,$$

then $f(t,\cdot)$ will be mean zero for all $t \geq 0$. In such case we obtain that $\|f\|_{L^2} \lesssim \|\nabla f\|_{L^2}$, or equivalently, the embedding $\dot{H}^1(\mathrm{D}) \hookrightarrow L^2(\mathrm{D})$ for functions f with mean zero and homogeneous Neumann boundary conditions. Assuming this, we obtain that the L^2 norm of solutions to the heat equation decays instantaneously and exponentially in time,

$$\|f(t,\cdot)\|_{L^2} \leq \|f(0,\cdot)\|_{L^2} e^{-\kappa t},$$

for all $t \geq 0$.

The type of inequality that we need for the *hypocoercivity estimate* in Section 5.6 is more involved since it features a weight in the norm of ∇f. In addition, we need to assume that D is convex. The weight in particular that we are interested in is a function distance to the boundary, that we will denote by

$$d_{\mathrm{D}}(x) = \mathrm{dist}(x, \partial\mathrm{D}) = \min\{|x-y| : y \in \partial\mathrm{D}\}.$$

Therefore this function has the property

$$\|\nabla d_{\mathrm{D}}\|_\infty \lesssim 1.$$

There are some very general results in the literature concerning weighted Poincaré inequalities. For instance let $\mathrm{D} \subset \mathbb{R}^d$ be convex and consider some parameters $\alpha \geq 0$, $\beta \in \mathbb{R}$, $1 \leq p \leq q < \infty$ such that

$$1 - \frac{d}{p} + \frac{d}{q} \geq 0, \quad 1 - \frac{d+\beta}{p} + \frac{d+\alpha}{q} \geq 0$$

then there holds

$$\|f - \mathrm{av}(f, d_{\mathrm{D}}^\alpha)\|_{L^q_\alpha} \leq C(\alpha,\beta,p,q)\|\nabla f\|_{L^p_\beta},$$

where

$$\operatorname{av}(f, d_{\mathrm{D}}^{\alpha}) = \left(\int_{\mathrm{D}} d_{\mathrm{D}}^{\alpha}(x)dx\right)^{-1} \int_{\mathrm{D}} f(x) d_{\mathrm{D}}^{\alpha}(x)dx,$$

and

$$\|f\|_{L^r_\gamma} = \left(\int_{\mathrm{D}} |f(x)|^r d_{\mathrm{D}}^{\gamma}(x)dx\right)^{1/r}.$$

See [23] for details about this result. For our case of interest, in Section 5.6, we deal with a two-dimensional disc $d = 2$ and hence it would suffice to choose $p = q = 2$, $\alpha = 0$ and $\beta = 2$, which precisely yields a weighted version of (A.10),

$$\left\| f - \fint_{\mathrm{D}} f dx \right\|_{L^2(\mathrm{D})} \lesssim \|d_{\mathrm{D}} \nabla f\|_{L^2(\mathrm{D})}.$$

The result in [23] is rather general and detailed, however, for the sake of completeness here we will give a proof of the weighted Poincaré inequality that we need to use in Section (A.10). First let us state a rather elementary result that is convenient to have for the proof of the main Proposition.

Lemma A.1. *Assume* $\mathrm{D} \subset \mathbb{R}^d$ *is bounded and* $f \in L^2(\mathrm{D})$*. Then there holds*

$$\inf_{c \in \mathbb{R}} \int_{\mathrm{D}} |f(x) - c|^2 dx = \int_{\mathrm{D}} \left| f(x) - \fint_{\mathrm{D}} f(z)dz \right|^2 dx.$$

Proof. The proof is just a consequence of a simple optimization problem. We consider the L^2 norm of $f(x) - c$ and expand the square,

$$\int_{\mathrm{D}} |f(x) - c|^2 dx = \|f\|^2_{L^2(\mathrm{D})} + c^2|\mathrm{D}| - 2c \int_{\mathrm{D}} f(x)dx.$$

Then if we optimize in c, we find the condition for the optimal $c \in \mathbb{R}$,

$$2c_{\text{optimal}}|D| - 2\int_{\mathrm{D}} f(x)dx = 0$$

that yields claim of the Lemma,

$$c_{\text{optimal}} = \fint_{\mathrm{D}} f(x)dx.$$

■

With this result in hand we can directly enunciate the weighted Poincaré inequality that we use in Section 5.6.

Proposition A.1 (Weighted Poincaré inequality)**.** *Let* $\mathrm{D} \subset \mathbb{R}^d$ *be a convex, bounded domain with smooth boundary. Let* $d_{\mathrm{D}}(x) = \mathrm{dist}(x, \partial\mathrm{D})$ *be the distance to the boundary, and let* $f \in H^1(\mathrm{D})$*. Then there holds*

$$\left\| f - \fint_{\mathrm{D}} f dx \right\|_{L^2(\mathrm{D})} \lesssim \|d_{\mathrm{D}} \nabla f\|_{L^2(\mathrm{D})}.$$

Proof. Consider $\varepsilon \in (0, \mathrm{diam}(\mathrm{D})/4)$ and let us define a cut-off function $\eta_\varepsilon \in C^\infty(\mathrm{D})$ of the form

$$\eta_\varepsilon(x) = \begin{cases} 1 & \text{if } d_{\mathrm{D}}(x) \geq 2\varepsilon \\ 0 & \text{if } d_{\mathrm{D}}(x) \leq \varepsilon \end{cases} \qquad \text{such that } \|\nabla \eta_\varepsilon\|_\infty \lesssim \frac{1}{\varepsilon}.$$

Before starting with the estimate, let us fix some convenient notation. For any $0 < \delta < \mathrm{diam}(\mathrm{D})$, we define

$$\mathrm{D}_\delta = \{x \in \mathrm{D} : \mathrm{dist}(x, \partial\mathrm{D}) > \delta\} \subset \mathrm{D}.$$

Given some number $c \in \mathbb{R}$ we can write

$$\int_{\mathrm{D}} |f(x) - c|^2 dx = \int_{\mathrm{D}} \eta_\varepsilon(x) |f(x) - c|^2 dx + \int_{\mathrm{D}} (1 - \eta_\varepsilon(x)) |f(x) - c|^2 dx.$$

For the first addend, given that $\eta_\varepsilon = 0$ if $d_{\mathrm{D}} \leq \varepsilon$, we can directly write

$$\int_{\mathrm{D}} \eta_\varepsilon(x) |f(x) - c|^2 dx \leq \int_{\mathrm{D}_\varepsilon} |f(x) - c|^2 dx$$

For the second addend, since we integrate now only near the boundary, we can smuggle in a factor $|\nabla d_{\mathrm{D}}|^2$ such that

$$\int_{\mathrm{D}} (1 - \eta_\varepsilon(x)) |f(x) - c|^2 dx \lesssim \int_{\mathrm{D}} (1 - \eta_\varepsilon(x)) |\nabla d_{\mathrm{D}}(x)|^2 |f(x) - c|^2 dx,$$

and integrating by parts,

$$
\begin{aligned}
&\int_{\mathrm{D}} (1-\eta_\varepsilon(x))|\nabla d_{\mathrm{D}}(x)|^2|f(x)-c|^2dx \\
&\qquad = -2\int_{\mathrm{D}} (1-\eta_\varepsilon(x))\nabla f(x)\cdot\nabla d_{\mathrm{D}}(x) d_{\mathrm{D}}(x)|f(x)-c|dx \\
&\qquad\quad - \int_{\mathrm{D}} (1-\eta_\varepsilon(x))\Delta d_{\mathrm{D}}(x) d_{\mathrm{D}}(x)|f(x)-c|^2dx \\
&\qquad\quad + 2\int_{\mathrm{D}} \nabla\eta_\varepsilon(x)\cdot\nabla d_{\mathrm{D}}(x) d_{\mathrm{D}}(x)|f(x)-c|^2dx \\
&\qquad = \mathrm{I}_\varepsilon + \mathrm{II}_\varepsilon + \mathrm{III}_\varepsilon\,.
\end{aligned}
$$

We will study these terms separately taking into account that the distance function and its derivatives are all bounded. On the one hand

$$
\begin{aligned}
\mathrm{I}_\varepsilon &\lesssim \int_{\mathrm{D}} d(x)|\nabla f(x)||f(x)-c|dx \\
&\lesssim \frac{1}{\varepsilon}\int_{\mathrm{D}} d(x)^2|\nabla f(x)|^2dx + \varepsilon\int_{\mathrm{D}} |f(x)-c|^2dx,
\end{aligned}
$$

On the other hand,

$$
\mathrm{II}_\varepsilon \lesssim \int_{\mathrm{D}\setminus\mathrm{D}_{2\varepsilon}} d_{\mathrm{D}}(x)|f(x)-c|^2 \lesssim \varepsilon\int_{\mathrm{D}} |f(x)-c|^2dx.
$$

Finally, the term involving the derivative of the cut-off function must be split if to regions, near the boundary and far from the boundary. We do the splitting in terms of the parameter ε^2 as follows,

$$
\begin{aligned}
\mathrm{III}_\varepsilon &\lesssim \int_{\mathrm{D}} d_{\mathrm{D}}(x)|\nabla\eta_\varepsilon(x)||f(x)-c|^2dx \\
&\lesssim \frac{1}{\varepsilon}\int_{\mathrm{D}\setminus\mathrm{D}_{\varepsilon^2}} d_{\mathrm{D}}(x)|f(x)-c|^2dx + \frac{1}{\varepsilon}\int_{\mathrm{D}_{\varepsilon^2}} d_{\mathrm{D}}(x)|f(x)-c|^2dx \\
&\lesssim \varepsilon\int_{\mathrm{D}} d_{\mathrm{D}}(x)|f(x)-c|^2dx + \frac{1}{\varepsilon}\int_{\mathrm{D}_{\varepsilon^2}} d_{\mathrm{D}}(x)|f(x)-c|^2dx.
\end{aligned}
$$

At this point there are only two terms left to be controlled and both have a very similar structure, namely

$$
\frac{1}{\varepsilon}\int_{\mathrm{D}_{\varepsilon^2}} d_{\mathrm{D}}(x)|f(x)-c|^2dx \quad \text{and} \int_{\mathrm{D}_\varepsilon} d_{\mathrm{D}}(x)|f(x)-c|^2dx.
$$

Assuming that $0 < \varepsilon < 1$ is small, there holds that

$$\int_{\mathrm{D}_\varepsilon} d_{\mathrm{D}}(x)|f(x)-c|^2 dx \leq \int_{\mathrm{D}_{\varepsilon^2}} d_{\mathrm{D}}(x)|f(x)-c|^2 dx \leq \frac{1}{\varepsilon}\int_{\mathrm{D}_{\varepsilon^2}} d_{\mathrm{D}}(x)|f(x)-c|^2 dx$$

and thus it will sufficient to find an appropriate bound for the latter. In order to do so we will use the standard Poincaré inequality from (A.10) in the domain

$$\mathrm{D}_{\varepsilon^2} = \{x \in \mathrm{D} : \mathrm{dist}(x, \partial\mathrm{D}) < \varepsilon^2\}.$$

There, by definition, we have that $d_{\mathrm{D}}(x) > \varepsilon^2$ and thus there yields

$$\frac{1}{\varepsilon}\int_{\mathrm{D}_{\varepsilon^2}} d_{\mathrm{D}}(x)|f(x)-c|^2 dx \lesssim \frac{1}{\varepsilon}\int_{\mathrm{D}_{\varepsilon^2}} |\nabla f(x)|^2 dx \lesssim \frac{1}{\varepsilon^5}\int_{\mathrm{D}} d_{\mathrm{D}}(x)^2|\nabla f(x)|^2 dx.$$

Notice that in order to apply (A.10) in the domain $\mathrm{D}_{\varepsilon^2}$ we use a specific value of $c \in \mathbb{R}$, in particular we need

$$c = \fint_{\mathrm{D}_{\varepsilon^2}} f(x)dx.$$

All in all, putting all the estimates together, we arrive to

$$(1-3\varepsilon)\int_{\mathrm{D}} |f(x)-c|^2 dx \lesssim \frac{1}{\varepsilon^5}\int_{\mathrm{D}} d_{\mathrm{D}}(x)^2|\nabla f(x)|^2 dx,$$

with a particular choice of $c \in \mathbb{R}$. As a consequence and choosing a suitable value for $\varepsilon > 0$ small enough, there holds

$$\inf_{c\in\mathbb{R}}\int_{\mathrm{D}} |f(x)-c|^2 dx \lesssim \int_{\mathrm{D}} d_{\mathrm{D}}(x)^2|\nabla f(x)|^2 dx,$$

and thus the claim of the Proposition follows thanks to Lemma A.1. ■

Bibliography

[1] G. Alberti, G. Crippa, and A. L. Mazzucato. Exponential self-similar mixing by incompressible flows. *J. Amer. Math. Soc.*, 32(2):445–490, 2019.

[2] S. Alesker, S. Dar, and V. Milman. A remarkable measure preserving diffeomorphism between two convex bodies in $\mathbb{R}^n$. *Geom. Dedicata*, 74(2):201–212, 1999.

[3] L. Ambrosio. Transport equation and Cauchy problem for BV vector fields. *Invent. Math.*, 158(2):227–260, 2004.

[4] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[5] L. Arnold. *Random dynamical systems.* Springer Monographs in Mathematics. Springer-Verlag, Berlin, 1998.

[6] J. Bedrossian, A. Blumenthal, and S. Punshon-Smith. Almost-sure enhanced dissipation and uniform-in-diffusivity exponential mixing for advection-diffusion by stochastic Navier-Stokes. *Probab. Theory Related Fields*, 179(3-4):777–834, 2021.

[7] J. Bedrossian, A. Blumenthal, and S. Punshon-Smith. Almost-sure exponential mixing of passive scalars by the stochastic Navier–Stokes equations. *Ann. Probab.*, 50(1):241 – 303, 2022.

[8] J. Bedrossian, A. Blumenthal, and S. Punshon-Smith. Lagrangian chaos and scalar advection in stochastic fluid mechanics. *J. Eur. Math. Soc. (JEMS)*, 24(6):1893–1990, 2022.

[9] J. Bedrossian, A. Blumenthal, and S. Punshon-Smith. A regularity method for lower bounds on the Lyapunov exponent for stochastic differential equations. *Invent. Math.*, 227(2):429–516, 2022.

[10] J. Bedrossian and M. Coti Zelati. Enhanced dissipation, hypoellipticity, and anomalous small noise inviscid limits in shear flows. *Arch. Ration. Mech. Anal.*, 224(3):1161–1204, 2017.

[11] S. Bianchini, M. Colombo, G. Crippa, and L. V. Spinolo. Optimality of integrability estimates for advection-diffusion equations. *NoDEA Nonlinear Differential Equations Appl.*, 24(4):Art. 33, 19, 2017.

[12] A. Blumenthal, M. Coti Zelati, and R. S. Gvalani. Exponential mixing for random dynamical systems and an example of Pierrehumbert. Preprint arXiv:2204.13651, 2022.

[13] V. I. Bogachev, N. V. Krylov, M. Röckner, and S. V. Shaposhnikov. *Fokker-Planck-Kolmogorov equations*, volume 207 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2015.

[14] P. Bonicatto, G. Ciampa, and G. Crippa. On the advection-diffusion equation with rough coefficients: weak solutions and vanishing viscosity. *arXiv preprint arXiv:2107.03659*, 2021.

[15] F. Bouchut and G. Crippa. Lagrangian flows for vector fields with gradient given by a singular integral. *J. Hyperbolic Differ. Equ.*, 10(2):235–282, 2013.

[16] F. Boyer. Analysis of the upwind finite volume method for general initial- and boundary-value transport problems. *IMA J. Numer. Anal.*, 32(4):1404–1439, 2012.

[17] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.

[18] Y. Brenier, F. Otto, and C. Seis. Upper bounds on coarsening rates in demixing binary viscous liquids. *SIAM J. Math. Anal.*, 43(1):114–134, 2011.

[19] A. Bressan. A lemma and a conjecture on the cost of rearrangements. *Rend. Sem. Mat. Univ. Padova*, 110:97–102, 2003.

[20] E. Bruè and Q.-H. Nguyen. Sharp regularity estimates for solutions of the continuity equation drifted by Sobolev vector fields. *Anal. PDE*, 14(8):2539–2559, 2021.

[21] J. A. Cañizo and S. Mischler. Harris-type results on geometric and subgeometric convergence to equilibrium for stochastic semigroups. *J. Funct. Anal.*, 284(7):Paper No. 109830, 46, 2023.

[22] A. Cheskidov and X. Luo. Nonuniqueness of weak solutions for the transport equation at critical space regularity. *Ann. PDE*, 7(1):Paper No. 2, 45, 2021.

[23] S.-K. Chua and R. L. Wheeden. Weighted Poincaré inequalities on convex domains. *Math. Res. Lett.*, 17(5):993–1011, 2010.

[24] G. Ciampa, G. Crippa, and S. Spirito. Smooth approximation is not a selection principle for the transport equation with rough vector field. *Calc. Var. Partial Differential Equations*, 59(1):Paper No. 13, 21, 2020.

[25] A. Clop, H. Jylhä, J. Mateu, and J. Orobitg. Well-posedness for the continuity equation for vector fields with suitable modulus of continuity. *J. Funct. Anal.*, 276(1):45–77, 2019.

[26] M. Colombo, M. Coti Zelati, and K. Widmayer. Mixing and diffusion for rough shear flows. *Ars Inven. Anal.*, pages Paper No. 2, 22, 2021.

[27] W. Cooperman. Exponential mixing by shear flows. Preprint arXiv:2206.14239, 2022.

[28] M. Coti Zelati. Stable mixing estimates in the infinite Péclet number limit. *J. Funct. Anal.*, 279(4):108562, 25, 2020.

[29] G. Crippa and C. De Lellis. Estimates and regularity results for the DiPerna-Lions flow. *J. Reine Angew. Math.*, 616:15–46, 2008.

[30] G. Crippa, R. Lucà, and C. Schulze. Polynomial mixing under a certain stationary Euler flow. *Phys. D*, 394:44–55, 2019.

[31] G. Crippa, C. Nobili, C. Seis, and S. Spirito. Eulerian and Lagrangian solutions to the continuity and Euler equations with L^1 vorticity. *SIAM J. Math. Anal.*, 49(5):3973–3998, 2017.

[32] G. Da Prato and J. Zabczyk. *Ergodicity for Infinite Dimensional Systems*. London Mathematical Society Lecture Note Series. Cambridge University Press, 1996.

[33] C. De Lellis. Ordinary differential equations with rough coefficients and the renormalization theorem of Ambrosio [after Ambrosio, DiPerna, Lions]. In *Séminaire Bourbaki - Volume 2006/2007 - Exposés 967-981*, number 317 in Astérisque. Société mathématique de France, 2008. talk:972.

[34] F. Delarue and F. Lagoutière. Probabilistic analysis of the upwind scheme for transport equations. *Arch. Ration. Mech. Anal.*, 199(1):229–268, 2011.

[35] F. Delarue, F. Lagoutière, and N. Vauchelet. Convergence order of upwind type schemes for transport equations with discontinuous coefficients. to appear in *J. Math. Pures Appl.*, 2017.

[36] F. Delarue, F. Lagoutière, and N. Vauchelet. Convergence analysis of upwind type schemes for the aggregation equation with pointy potential. *Annales Henri Lebesgue*, 3:217–260, 2020.

[37] R. J. DiPerna and P.-L. Lions. Ordinary differential equations, transport theory and Sobolev spaces. *Invent. Math.*, 98(3):511–547, 1989.

[38] C. R. Doering and J.-L. Thiffeault. Multiscale mixing efficiencies for steady sources. *Phys. Rev. E*, 74:025301, Aug 2006.

[39] J. L. Doob. Asymptotic properties of Markoff transition probabilities. *Transactions of the American Mathematical Society*, 63(3):393–421, 1948.

[40] T. M. Elgindi and A. Zlatoš. Universal mixers in all dimensions. *Advances in Mathematics*, 356:106807, Nov. 2019.

[41] L. C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010.

[42] L. C. Evans and R. F. Gariepy. *Measure theory and fine properties of functions*. Textbooks in Mathematics. CRC Press, Boca Raton, FL, revised edition, 2015.

[43] R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. In *Handbook of numerical analysis*, Handb. Numer. Anal., VII, pages 713–1020. North-Holland, Amsterdam, 2000.

[44] M. Fathi and D. Mikulincer. Stability estimates for invariant measures of diffusion processes, with applications to stability of moment measures and stein kernels. *Preprint arXiv:2010.14178*, 2020.

[45] A. Figalli. Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients. *J. Funct. Anal.*, 254(1):109–153, 2008.

[46] N. Fournier and B. Perthame. Monge-Kantorovich distance for PDEs: the coupling method. *EMS Surv. Math. Sci.*, 7(1):1–31, 2021.

[47] A. Friedman. *Partial differential equations of parabolic type*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1964.

[48] W. Gangbo and R. J. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996.

[49] D. S. Grebenkov and B.-T. Nguyen. Geometrical structure of laplacian eigenfunctions. *SIAM Review*, 55(4):601–667, 2013.

[50] M. Hairer. Ergodic properties of Markov processes. Lecture notes, 2018.

[51] M. Hairer. Convergence of Markov processes. Lecture notes, 2021.

[52] M. Hairer and J. C. Mattingly. Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI*, volume 63 of *Progr. Probab.*, pages 109–117. Birkhäuser/Springer Basel AG, Basel, 2011.

[53] L. Hörmander. Hypoelliptic second order differential equations. *Acta Math.*, 119:147–171, 1967.

[54] G. Iyer, A. Kiselev, and X. Xu. Lower bounds on the mix norm of passive scalars advected by incompressible enstrophy-constrained flows. *Nonlinearity*, 27(5):973–985, 2014.

[55] P.-E. Jabin. Differential equations with singular fields. *J. Math. Pures Appl. (9)*, 94(6):597–621, 2010.

[56] N. N. Kuznetsov. The accuracy of certain approximate methods for the computation of weak solutions of a first order quasilinear equation. *Ž. Vyčisl. Mat. i Mat. Fiz.*, 16(6):1489–1502, 1627, 1976.

[57] O. A. Ladyženskaja. Uniqueness and smoothness of generalized solutions of Navier-Stokes equations. *Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)*, 5:169–185, 1967.

[58] O. A. Ladyženskaja, V. A. Solonnikov, and N. N. Ural'ceva. *Linear and quasilinear equations of parabolic type.* Translated from the Russian by S. Smith. Translations of Mathematical Monographs, Vol. 23. American Mathematical Society, Providence, R.I., 1968.

[59] C. Le Bris and P.-L. Lions. Existence and uniqueness of solutions to Fokker-Planck type equations with irregular coefficients. *Comm. Partial Differential Equations*, 33(7-9):1272–1317, 2008.

[60] C. Le Bris and P.-L. Lions. *Parabolic Equations with Irregular Data and Related Issues.* De Gruyter, 2019.

[61] H. Li and D. Luo. Quantitative stability estimates for Fokker-Planck equations. *J. Math. Pures Appl. (9)*, 122:125–163, 2019.

[62] E. Lunasin, Z. Lin, A. Novikov, A. Mazzucato, and C. R. Doering. Optimal mixing and optimal stirring for fixed energy, fixed power, or fixed palenstrophy flows. *J. Math. Phys.*, 53(11):115611, 15, 2012.

[63] G. Mathew, I. Mezić, and L. Petzold. A multiscale measure for mixing. *Phys. D*, 211(1-2):23–46, 2005.

[64] R. J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80(2):309–323, 1995.

[65] B. Merlet. L^∞- and L^2-error estimates for a finite volume approximation of linear advection. *SIAM J. Numer. Anal.*, 46(1):124–150, 2007/08.

[66] D. Meyer and C. Seis. Propagation of regularity for transport equations. A Littlewood-Paley approach. Preprint arXiv:2203.10860, 2023.

[67] S. P. Meyn and R. L. Tweedie. Stability of Markovian processes II: Continuous-time processes and sampled chains. *Advances in Applied Probability*, 25(3):487–517, 1993.

[68] C. J. Miles and C. R. Doering. Diffusion-limited mixing by incompressible flows. *Nonlinearity*, 31(5):2346, apr 2018.

[69] S. Modena and G. Sattig. Convex integration solutions to the transport equation with full dimensional concentration. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 37(5):1075–1108, 2020.

[70] S. Modena and L. Székelyhidi, Jr. Non-uniqueness for the transport equation with Sobolev vector fields. *Ann. PDE*, 4(2):Paper No. 18, 38, 2018.

[71] V. Navarro-Fernández and A. Schlichting. Error estimates for a finite volume scheme for advection-diffusion equations with rough coefficients. *ESAIM: M2AN*, 57(4):2131–2158, 2023.

[72] V. Navarro-Fernández, A. Schlichting, and C. Seis. Optimal stability estimates and a new uniqueness result for advection-diffusion equations. *Pure Appl. Anal.*, 4(3):571–596, 2022.

[73] H. J. Nussenzveig Lopes, C. Seis, and E. Wiedemann. On the vanishing viscosity limit for 2D incompressible flows with unbounded vorticity. *Preprint arXiv:2007.01091*, 2020.

[74] J. M. Ottino. Mixing, chaotic advection, and turbulence. *Annual Review of Fluid Mechanics*, 22(1):207–254, 1990.

[75] F. Otto, C. Seis, and D. Slepčev. Crossover of the coarsening rates in demixing of binary viscous liquids. *Commun. Math. Sci.*, 11(2):441–464, 2013.

[76] A. Pilipenko. *An introduction to stochastic differential equations with reflection.* Lectures in pure and applied mathematics. Universitätsverlag Potsdam, 2014.

[77] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, third edition, 1999.

[78] A. A. Samarskiĭ. On monotone difference schemes for elliptic and parabolic equations in the case of a non-selfadjoint elliptic operator. *Ž. Vyčisl. Mat i Mat. Fiz.*, 5:548–551, 1965.

[79] A. Schlichting and C. Seis. Convergence rates for upwind schemes with rough coefficients. *SIAM J. Numer. Anal.*, 55(2):812–840, 2017.

[80] A. Schlichting and C. Seis. Analysis of the implicit upwind finite volume scheme with rough coefficients. *Numer. Math.*, 139(1):155–186, 2018.

[81] A. Schlichting and C. Seis. The Scharfetter-Gummel scheme for aggregation-diffusion equations. *IMA J. Numer. Anal.*, 42(3):2361–2402, 2022.

[82] C. Seis. Maximal mixing by incompressible fluid flows. *Nonlinearity*, 26(12):3279–3289, 2013.

[83] C. Seis. A quantitative theory for the continuity equation. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 34(7):1837–1850, 2017.

[84] C. Seis. Optimal stability estimates for continuity equations. *Proc. Roy. Soc. Edinburgh Sect. A*, 148(6):1279–1296, 2018.

[85] C. Seis. On the Littlewood-Paley spectrum for passive scalar transport equations. *J. Nonlinear Sci.*, 30(2):645–656, 2020.

[86] B. I. Shraiman and E. D. Siggia. Scalar turbulence. *Nature*, 405(6787):639–646, June 2000.

[87] E. M. Stein. *Singular integrals and differentiability properties of functions.* Princeton Mathematical Series, No. 30. Princeton University Press, 1970.

[88] D. W. Stroock and S. R. S. Varadhan. Diffusion processes with continuous coefficients. I. *Comm. Pure Appl. Math.*, 22:345–400, 1969.

[89] D. W. Stroock and S. R. S. Varadhan. Diffusion processes with continuous coefficients. II. *Comm. Pure Appl. Math.*, 22:479–530, 1969.

[90] D. W. Stroock and S. R. S. Varadhan. Diffusion processes with boundary conditions. *Comm. Pure Appl. Math.*, 24:147–225, 1971.

[91] J.-L. Thiffeault. Using multiscale norms to quantify mixing and transport. *Nonlinearity*, 25(2):R1–R44, 2012.

[92] D. Trevisan. Well-posedness of multidimensional diffusion processes with weakly differentiable coefficients. *Electron. J. Probab.*, 21:Paper No. 22, 41, 2016.

[93] R. Vanselow. Relations between FEM and FVM applied to the Poisson equation. *Computing*, 57(2):93–104, 1996.

[94] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[95] C. Villani. Hypocoercivity. *Mem. Amer. Math. Soc.*, 202(950):iv+141, 2009.

[96] Y. Yao and A. Zlatoš. Mixing and un-mixing by incompressible flows. *J. Eur. Math. Soc. (JEMS)*, 19(7):1911–1948, 2017.

[97] X. Zhang. Well-posedness and large deviation for degenerate SDEs with Sobolev coefficients. *Rev. Mat. Iberoam.*, 29(1):25–52, 2013.